



DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITE DE COOPERATION EN MATIÈRE DE BREVETS (PCT)

| | | | |
|--|--|--|--|
| (51) Classification internationale des brevets ⁶ : C12N 15/52, C12P 19/62, C12Q 1/68, C12N 1/20 // (C12N 1/20, C12R 1:01) | | A2 | (11) Numéro de publication internationale: WO 99/05283 |
| | | | (43) Date de publication internationale: 4 février 1999 (04.02.99) |
| (21) Numéro de la demande internationale: PCT/FR98/01593 | | Guillermo Estrada, 2-Bajo Izquierda. E-33060 Oviedo (ES). | |
| (22) Date de dépôt international: 21 juillet 1998 (21.07.98) | | (74) Mandataire: VIEILLEFOSSE, Jean, Claude; Hoechst Marion Roussel, 102, route de Noisy, F-93235 Romainville Cedex (FR). | |
| (30) Données relatives à la priorité: 97/09458 25 juillet 1997 (25.07.97) FR 98/07411 12 juin 1998 (12.06.98) FR | | (81) Etats désignés: BR, CA, JP, MX, TR, US, brevet européen (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). | |
| (71) Déposant (pour tous les Etats désignés sauf US): HOECHST MARION ROUSSEL [FR/FR]; 1, terrasse Bellini, F-92800 Puteaux (FR). | | Publiée. Sans rapport de recherche internationale, sera republiée dès réception de ce rapport. | |
| (72) Inventeurs; et (75) Inventeurs/Déposants (US seulement): FROMENTIN, Claude [FR/FR]; 16, rue de Flandres, F-75019 Paris (FR). MICHEL, Jean-Marc [FR/FR]; 22, rue des Domeliers, F-60200 Compiègne (FR). RAYNAL, Marie-Cécile [FR/FR]; 117, avenue de Choisy, F-75013 Paris (FR). SALAH-BEY, Khadidja [DZ/FR]; Appartement 2042, 100, boulevard Masséna, F-75013 Paris (FR). CORTES, Jesus [MX/GB]; 26 Cambanks, Union Lane, Cambridge CB4 1PZ (GB). GAISSER, Sabine [DE/GB]; 37 Gwydir Street, Cambridge CB1 2LG (GB). LEADLAY, Peter [GB/GB]; 17 Clarendon Road, Cambridge CB2 2BH (GB). MENDEZ, Carmen [ES/ES]; Calle Marcelino Fernandez 7, 2°B, E-33010 Oviedo (ES). SALAS, Jose, A. [ES/ES]; Calle | | | |
| (54) Title: BIOSYNTHESIS GENES AND TRANSFER OF 6-DESOXY-HEXOSES IN SACCHAROPOLYSPORA ERYTHRAEA AND IN STREPTOMYCES ANTIBIOTICUS AND THEIR USE | | | |
| (54) Titre: GENES DE BIOSYNTHESE ET DE TRANSFERT DES 6-DESOXYHEXOSES CHEZ SACCHAROPOLYSPORA ERYTHRAEA ET CHEZ STREPTOMYCES ANTIBIOTICUS ET LEUR UTILISATION | | | |
| (57) Abstract | | | |
| <p>The invention concerns the isolated DNA sequence represented in figure 2 (SEQ ID No. 1) corresponding to the <i>eryG-eryAIII</i> region of the cluster of the erythromycin biosynthesis genes and the isolated DNA sequence represented in figure 3 (SEQ ID No. 6) corresponding to the <i>eryAI-eryK</i> region of the cluster of the erythromycin biosynthesis genes. The invention also concerns the isolated DNA sequence represented in figure 22 (SEQ ID No. 15 sequence) corresponding to a region of the oleandomycin biosynthesis genes (SEQ ID No. 15 sequence).</p> | | | |
| (57) Abrégé | | | |
| <p>L'invention a pour objet la séquence d'ADN isolée représentée à la figure 2 (SEQ ID No. 1) correspondant à la région <i>eryG-eryAIII</i> du cluster de gènes de la biosynthèse de l'érythromycine et la séquence d'ADN isolée représentée à la figure 3 (SEQ ID No. 6) correspondant à la région <i>eryAI-eryK</i> du cluster de gènes de la biosynthèse de l'érythromycine, et a pour objet la séquence d'ADN isolée représentée à la figure 22 (séquence de SEQ ID No. 15) correspondant à une région du cluster de gènes de la biosynthèse de l'oléandomycine (séquence de SEQ ID No. 15).</p> | | | |

BEST AVAILABLE COPY

UNIQUEMENT A TITRE D'INFORMATION

Codes utilisés pour identifier les Etats parties au PCT, sur les pages de couverture des brochures publiant des demandes internationales en vertu du PCT.

| | | | | | | | |
|----|---------------------------|----|---|----|--|----|-----------------------|
| AL | Albanie | ES | Espagne | LS | Lesotho | SI | Slovénie |
| AM | Arménie | FI | Finlande | LT | Lituanie | SK | Slovaquie |
| AT | Autriche | FR | France | LU | Luxembourg | SN | Sénégal |
| AU | Australie | GA | Gabon | LV | Lettonie | SZ | Swaziland |
| AZ | Azerbaïdjan | GB | Royaume-Uni | MC | Monaco | TD | Tchad |
| BA | Bosnie-Herzégovine | GE | Géorgie | MD | République de Moldova | TG | Togo |
| BB | Barbade | GH | Ghana | MG | Madagascar | TJ | Tadjikistan |
| BE | Belgique | GN | Guinée | MK | Ex-République yougoslave de Macédoine | TM | Turkménistan |
| BF | Burkina Faso | GR | Grèce | ML | Mali | TR | Turquie |
| BG | Bulgarie | HU | Hongrie | MN | Mongolie | TT | Trinité-et-Tobago |
| BJ | Bénin | IE | Irlande | MR | Mauritanie | UA | Ukraine |
| BR | Bésil | IL | Israël | MW | Malawi | UG | Ouganda |
| BY | Biélorus | IS | Islande | MX | Mexique | US | Etats-Unis d'Amérique |
| CA | Canada | IT | Italie | NE | Niger | UZ | Ouzbékistan |
| CF | République centrafricaine | JP | Japon | NL | Pays-Bas | VN | Viet Nam |
| CG | Congo | KE | Kenya | NO | Norvège | YU | Yougoslavie |
| CH | Suisse | KG | Kirghizistan | NZ | Nouvelle-Zélande | ZW | Zimbabwe |
| CI | Côte d'Ivoire | KP | République populaire démocratique de Corée | PL | Pologne | | |
| CM | Cameroun | KR | République de Corée | PT | Portugal | | |
| CN | Chine | KZ | Kazakhstan | RO | Roumanie | | |
| CU | Cuba | LC | Sainte-Lucie | RU | Fédération de Russie | | |
| CZ | République tchèque | LI | Liechtenstein | SD | Soudan | | |
| DE | Allemagne | LK | Sri Lanka | SE | Suède | | |
| DK | Danemark | LR | Libéria | SG | Singapour | | |
| EE | Estonie | | | | | | |

Gènes de biosynthèse et de transfert des 6-désoxyhexoses chez *Saccharopolyspora erythraea* et chez *Streptomyces antibioticus* et leur utilisation.

- 5 La présente invention décrit des gènes impliqués dans la biosynthèse et le transfert des 6-désoxyhexoses chez *Saccharopolyspora erythraea* et leur utilisation dans la production d'analogues de l'érythromycine par manipulation génétique.
- 10 L'érythromycine A est un antibiotique macrolide cliniquement important produit par la bactérie gram-positive *Sac. erythraea*. Les gènes de la biosynthèse de l'érythromycine sont organisés en un cluster de gènes *ery* qui inclut aussi le gène d'auto-résistance à l'érythromycine *ermE*.
- 15 Le cluster *ery* contient les trois grands gènes *eryAI*, *eryAII* et *eryAIII* (locus *eryA*) codant pour trois polypeptides composant la polykétide synthétase (dénommée PKS) flanqués par deux régions comprenant les gènes impliqués dans les stades ultérieurs de conversion du noyau lactone
- 20 (6-désoxyérythronolide B) en érythromycine A.
- Pendant le processus de biosynthèse de l'érythromycine A représenté à la figure 1, la biosynthèse des 6-désoxyhexoses comprend l'ensemble des réactions enzymatiques conduisant du glucose-1-phosphate au sucre activé final dTDP-L-mycarose ou
- 25 dTDP-D-désosamine. Le dTDP-L-mycarose ou la dTDP-D-désosamine ainsi produits sont ensuite utilisés comme substrats pour le transfert des deux désoxyhexoses sur le noyau lactone. La formation de l'érythromycine requiert l'attachement du mycarose via l'hydroxyle en position C-3 du noyau lactone et
- 30 l'attachement de la désosamine via l'hydroxyle en position C-5. L'ensemble des gènes *eryB* impliqués dans la biosynthèse ou le transfert du mycarose et l'ensemble des gènes *eryC* impliqués dans la biosynthèse ou le transfert de la désosamine n'ont pas encore été clairement identifiés.
- 35 Le cluster *ery* d'une longueur de 56 kb comprend 21 phases ouvertes de lecture ou "open reading frames" (ORFs) dont la numérotation a été établie par Haydock et al. (1991) et Donadio et al. (1993). Le locus *eryA* comprend les ORFs 10,

11 et 12.

Des travaux précoces d'interruption ou de remplacement de gène dans la partie gauche du cluster ery a permis une première identification du gène eryCI (ORF1) (Dhillon et al., 1989), puis du gène eryBI (ORF2), du locus eryH (ORFs 3, 4 et 5) dont l'inactivation conduit à la production de 6-désoxyérythronolide B, d'un locus eryBII (ORFs 7 et 8) et le gène eryCII (Weber et al., 1990).

Parmi les activités enzymatiques impliquées dans les modifications ultérieures du noyau lactone ont été identifiées le gène eryF (ORF4) responsable de l'hydroxylation en C6 (Weber et al., 1991) et le gène eryK (ORF20) responsable de l'hydroxylation en C12 (Stassi et al., 1993). D'autre part, le gène eryG (ORF6) responsable de la O-méthylation du mycarose en cladinose (position 3"OH) a été identifié (Weber et al., 1989). L'érythromycine A est ainsi formée via l'érythromycine B ou l'érythromycine C à partir de l'érythromycine D selon le schéma proposé (figure 1).

La caractérisation fonctionnelle des gènes eryB et eryC situés sur la partie droite de cluster ery (ORFs 13 à 19) n'a pas encore été établie de façon précise, malgré les informations parcellaires communiquées dans différents articles de revues (Donadio et al., 1993 ; Liu et Thorson, 1994 ; Katz et Donadio, 1995).

En raison de l'intérêt commercial des antibiotiques macrolides, l'obtention de nouveaux dérivés, notamment l'obtention d'analogues de l'érythromycine ayant des propriétés avantageuses, est intensivement recherchée. Les modifications peuvent être désirées dans la partie aglycone (macrolactone) ou/et dans son hydroxylation secondaire ainsi que dans la partie sucre (cladinose et/ou désosamine) de l'érythromycine.

Les méthodes courantes telles que les modifications chimiques sont difficiles et limitées vis-à-vis du type de produit que l'on peut obtenir à partir de l'érythromycine. Par exemple, Sakakibara et al. (1984) passent en revue des modifications chimiques réalisées à partir de l'érythromycine A ou B, aussi bien dans la partie sucre que dans la macro-

lactone.

Des modifications de la macrolactone de l'érythromycine A par manipulation génétique du microorganisme *Sac. erythraea* ont été décrites dans la demande de brevet internationale WO 93/13663 ainsi que l'obtention de nouvelles molécules polykétides par altérations génétiques spécifiques du locus *eryA* du chromosome codant pour la PKS. Par exemple la 7-hydroxyérythromycine A, la 6-désoxy-7-hydroxyérythromycine A ou le 3-oxo-3-désoxy-5-désoaminyl-érythronolide A ont été ainsi obtenus.

La présente invention concerne la caractérisation fonctionnelle de dix gènes de *Sac. erythraea* impliqués dans la biosynthèse ou l'attachement du mycarose et de la désosamine (*eryBII*, *eryCIII* et *eryCII* situés en aval du locus *eryA* et *eryBIV*, *eryBV*, *eryCVI*, *eryBVI*, *eryCIV*, *eryCV* et *eryBVII* situés en amont), leur utilisation dans la production d'analogues de l'érythromycine ainsi qu'un procédé de préparation de ceux-ci.

La présente invention a donc pour objet une séquence d'ADN simple ou double brin isolée, représentée à la figure 2 (séquence directe et complémentaire de SEQ ID N° 1) correspondant à la région *eryG-eryAIII* du cluster de gènes de la biosynthèse de l'érythromycine et a particulièrement pour objet une séquence d'ADN ci-dessus comprenant :

- 25 - la séquence *eryBII* correspondant à l'ORF7 (séquence complémentaire de SEQ ID N° 1 du nucléotide 48 au nucléotide 1046) et codant pour une dTDP-4-céto-L-6-désoxyhexose 2,3-réductase,
- la séquence *eryCIII* correspondant à l'ORF8 (séquence complémentaire de SEQ ID N° 1 du nucléotide 1046 au nucléotide 2308) et codant pour une désosaminyltransférase et
- 30 - la séquence *eryCII* correspondant à l'ORF9 (séquence complémentaire de SEQ ID N° 1 du nucléotide 2322 au nucléotide 3404) et codant pour une dTDP-4-céto-D-6-désoxyhexose
- 35 3,4-isomérase.

La séquence d'ADN ci-dessus montrée à la figure 2 est une séquence d'ADN génomique qui peut être obtenue par exemple par sous-clonage de fragments de restriction d'un

fragment d'ADN génomique de *Sac. erythraea*, selon des conditions opératoires dont une description détaillée est donnée plus loin.

L'invention a plus particulièrement pour objet une
5 séquence d'ADN isolée représentée à la figure 2 choisie parmi la séquence eryBII correspondant à l'ORF7 (séquence complémentaire de SEQ ID N° 1 du nucléotide 48 au nucléotide 1046), la séquence eryCIII correspondant à l'ORF8 (séquence complémentaire de SEQ ID N° 1 du nucléotide 1046 au nucléotide
10 2308) ou la séquence eryCII correspondant à l'ORF9 (séquence complémentaire de SEQ ID N° 1 du nucléotide 2322 au nucléotide 3404) et les séquences qui hybrident et/ou présentent des homologies significatives avec cette séquence ou des fragments de celle-ci et ayant la même fonction.

15 L'invention a tout particulièrement pour objet la séquence d'ADN isolée eryCIII représentée à la figure 2 correspondant à l'ORF8 (séquence complémentaire de SEQ ID N° 1 du nucléotide 1046 au nucléotide 2308 = séquence complémentaire de SEQ ID N° 4) et codant pour une désosaminyl-
20 transférase.

La séquence eryBII correspondant à l'ORF7 code pour un polypeptide ayant 333 acides aminés (séquence de SEQ ID N° 2), la séquence eryCIII correspondant à l'ORF8 code pour un polypeptide ayant 421 acides aminés (séquence de SEQ ID
25 N° 5) et la séquence eryCII correspondant à l'ORF9 code pour un polypeptide ayant 361 acides aminés (séquence de SEQ ID N° 3).

La mise en évidence des activités enzymatiques respectives indiquées ci-dessus a été réalisée par introduction d'une
30 délétion interne au gène correspondant telle qu'illustrée plus loin dans la partie expérimentale.

Par séquences qui hybrident et ayant la même fonction, on inclut les séquences d'ADN qui hybrident avec l'une des séquences d'ADN ci-dessus sous des conditions standards de
35 stringence élevée ou moyenne décrites par Sambrook et al. (1989) et qui codent pour une protéine ayant la même fonction enzymatique. Par même fonction enzymatique, on entend une activité enzymatique donnée sur des substrats de même nature,

par exemple un dTDP-6-désoxyhexose ou une macrolactone nue ou glycosylée. Les conditions de forte stringence comprennent par exemple une hybridation à 65°C pendant 18 heures dans une solution 5 x SSPE, 10 x Denhardt, 100 µg/ml DNase, 1 % SDS
5 suivie de 2 lavages pendant 20 minutes avec une solution 2 x SSC, 0,05 % SDS à 65°C suivis d'un dernier lavage pendant 45 minutes dans une solution 0,1 x SSC, 0,1 % SDS à 65°C. Les conditions de stringence moyenne comprennent par exemple un dernier lavage pendant 20 minutes dans une solution 0,2 x
10 SSC, 0,1 % SDS à 65°C.

Par séquences qui présentent des homologues significatives et ayant la même fonction, on inclut les séquences ayant une identité de séquence nucléotidique d'au moins 60 % avec l'une des séquences ADN ci-dessus et qui codent pour une
15 protéine ayant la même fonction enzymatique.

L'invention a aussi pour objet un polypeptide codé par l'une des séquences d'ADN ci-dessus et a spécialement pour objet un polypeptide correspondant à une ORF représentée à la figure 2, choisie parmi l'ORF7 (ayant la séquence de SEQ ID
20 N° 2), l'ORF8 (ayant la séquence de SEQ ID N° 5) ou l'ORF9 (ayant la séquence de SEQ ID N° 3) et les analogues de ce polypeptide.

Par analogues, on inclut les peptides ayant une séquence en acides aminés modifiée par substitution, délétion ou
25 addition d'un ou plusieurs acides aminés pour autant que ces produits conservent la même fonction enzymatique. Les séquences modifiées peuvent être par exemple préparées en utilisant la technique de mutagenèse dirigée connue de l'homme du métier.

30 L'invention a plus spécialement pour objet le polypeptide correspondant à l'ORF 8 représentée à la figure 2 (ayant la séquence de SEQ ID N° 5) et ayant une activité désosaminyltransférase, dénommé EryCIII.

L'invention décrit une protéine recombinante EryCIII de
35 *Sac. erythraea* obtenue par expression dans une cellule hôte selon les méthodes connues de génie génétique et de culture cellulaire.

L'obtention de la protéine recombinante purifiée a

permis de confirmer la caractérisation de la fonction glycosyltransférase associée au produit du gène *eryCIII* dans un test *in vitro* qui met en évidence le transfert du sucre activé dTDP-D-désosamine sur le noyau lactone.

- 5 L'invention a aussi pour objet la thymidine 5'-(tri-hydrogène diphosphate), P'-[3,4,6-tridésoxy-3-(diméthylamino)-D-.xylo.-hexopyranosyl] ester (dTDP-D-désosamine) et les sels d'addition avec les bases, dont un exemple de préparation est décrit plus loin dans la partie expérimentale.
- 10 L'invention a aussi pour objet une séquence d'ADN isolée représentée à la figure 3 (séquence de SEQ ID N° 6) correspondant à la région *eryAI-eryK* du cluster de gènes de la biosynthèse de l'érythromycine et a particulièrement pour objet une séquence d'ADN ci-dessus comprenant :
- 15 - la séquence *eryBIV* correspondant à l'ORF13 (séquence de SEQ ID N° 6 du nucléotide 242 au nucléotide 1207) et codant pour une dTDP-4-céto-L-6-désoxyhexose 4-réductase,
- la séquence *eryBV* correspondant à l'ORF14 (séquence de SEQ ID N° 6 du nucléotide 1210 au nucléotide 2454) et codant pour
20 une mycarosyltransférase,
- la séquence *eryCVI* correspondant à l'ORF15 (séquence de SEQ ID N° 6 du nucléotide 2510 au nucléotide 3220) et codant pour une dTDP-D-6-désoxyhexose 3-N-méthyltransférase,
- la séquence *eryBVI* correspondant à l'ORF16 (séquence de SEQ
25 ID N° 6 du nucléotide 3308 au nucléotide 4837) et codant pour une dTDP-4-céto-L-6-désoxyhexose 2,3-déshydratase,
- la séquence *eryCIV* correspondant à l'ORF17 (séquence de SEQ ID N° 6 du nucléotide 4837 au nucléotide 6039) et codant pour une dTDP-D-6-désoxyhexose 3,4-déshydratase,
30 - la séquence *eryCV* correspondant à l'ORF18 (séquence de SEQ ID N° 6 du nucléotide 6080 au nucléotide 7546) et codant pour une dTDP-D-4,6-didésoxyhexose 3,4-réductase et
- la séquence *eryBVII* correspondant à l'ORF19 (séquence de SEQ ID N° 6 du nucléotide 7578 au nucléotide 8156) et codant
35 pour une dTDP-4-céto-D-6-désoxyhexose 3,5 épimérase.

La séquence d'ADN ci-dessus montrée à la figure 3 est une séquence d'ADN génomique qui peut être obtenue, par exemple, par sous-clonage de fragments de restriction de

cosmides contenant une banque d'ADN génomique de *Sac. erythraea*, selon des conditions opératoires dont une description détaillée est donnée plus loin.

L'invention a plus particulièrement pour objet une
5 séquence d'ADN isolée représentée à la figure 3 choisie parmi la séquence eryBIV correspondant à l'ORF13 (séquence de SEQ ID N° 6 du nucléotide 242 au nucléotide 1207), la séquence eryBV correspondant à l'ORF14 (séquence de SEQ ID N° 6 du nucléotide 1210 au nucléotide 2454), la séquence eryCVI
10 correspondant à l'ORF15 (séquence de SEQ ID N° 6 du nucléotide 2510 au nucléotide 3220), la séquence eryBVI correspondant à l'ORF16 (séquence de SEQ ID N° 6 du nucléotide 3308 au nucléotide 4837), la séquence eryCIV correspondant à l'ORF17 (séquence de SEQ ID N° 6 du nucléotide 4837 au nucléotide
15 6039), la séquence eryCV correspondant à l'ORF18 (séquence de SEQ ID N° 6 du nucléotide 6080 au nucléotide 7546) ou la séquence eryBVII correspondant à l'ORF19 (séquence de SEQ ID N° 6 du nucléotide 7578 au nucléotide 8156) et les séquences qui hybrident et/ou présentent des homologues significatives
20 avec cette séquence ou des fragments de celle-ci et ayant la même fonction.

L'invention a tout particulièrement pour objet la séquence d'ADN isolée eryBV représentée à la figure 3 correspondant à l'ORF14 (séquence de SEQ ID N° 6 du nucléotide
25 1210 au nucléotide 2454) et codant pour une mycarosyl-transférase.

La séquence eryBIV correspondant à l'ORF13 code pour un polypeptide ayant 322 acides aminés (SEQ ID N° 7), la séquence eryBV correspondant à l'ORF14 code pour un polypeptide
30 ayant 415 acides aminés (SEQ ID N° 8), la séquence eryCVI correspondant à l'ORF15 code pour un polypeptide ayant 237 acides aminés (SEQ ID N° 9), la séquence eryBVI correspondant à l'ORF16 code pour un polypeptide ayant 510 acides aminés (SEQ ID N° 10), la séquence eryCIV correspondant à
35 l'ORF17 code pour un polypeptide ayant 401 acides aminés (SEQ ID N° 14), la séquence eryCV correspondant à l'ORF18 code pour un polypeptide ayant 489 acides aminés (SEQ ID N° 11) et la séquence eryBVII correspondant à l'ORF19 code pour un

polypeptide ayant 193 acides aminés (SEQ ID N° 12).

La mise en évidence des activités enzymatiques respectives indiquées ci-dessus a été réalisée par introduction d'une délétion interne au gène correspondant telle

5 qu'illustrée plus loin dans la partie expérimentale.

Les séquences d'ADN qui hybrident ainsi que les séquences d'ADN qui présentent des homologies significatives et ayant la même fonction ont la même signification que celle indiquée précédemment.

10 L'invention a aussi pour objet un polypeptide codé par l'une des séquences d'ADN ci-dessus et a spécialement pour objet un polypeptide correspondant à une ORF représentée à la figure 3, choisie parmi l'ORF13 (ayant la séquence de SEQ ID N° 7), l'ORF14 (ayant la séquence de SEQ ID N° 8), l'ORF15
15 (ayant la séquence de SEQ ID N° 9), l'ORF16 (ayant la séquence de SEQ ID N° 10), l'ORF17 (ayant la séquence de SEQ ID N° 14), l'ORF18 (ayant la séquence de SEQ ID N° 11) ou l'ORF19 (ayant la séquence de SEQ ID N° 12) et les analogues de ce peptide.

20 Les analogues du polypeptide ont la même signification que celle indiquée précédemment.

L'invention a plus spécialement pour objet le polypeptide correspondant à l'ORF14 représentée à la figure 3 (ayant la séquence de SEQ ID N° 8) et ayant une activité
25 mycarosyltransférase, dénommé EryBV.

La connaissance de chaque séquence d'ADN eryB ou eryC de l'invention indiquée ci-dessus et montrée à la figure 2 ou à la figure 3 permet de reproduire la présente invention par exemple par des méthodes connues de synthèse chimique ou par
30 criblage d'une banque génomique à l'aide de sondes d'oligonucléotides de synthèse par les techniques connues d'hybridation ou par amplification par PCR.

Les polypeptides de l'invention peuvent être obtenus par les méthodes connues, par exemple par synthèse chimique ou
35 par la méthodologie de l'ADN recombinant par expression dans une cellule hôte procaryote ou eucaryote.

Un autre objet de l'invention concerne l'utilisation d'au moins l'une des séquences d'ADN choisie parmi les

séquences eryBII (séquence complémentaire de SEQ ID N° 1 du nucléotide 48 au nucléotide 1046), eryCIII (séquence complémentaire de SEQ ID N° 1 du nucléotide 1046 au nucléotide 2308) ou eryCII (séquence complémentaire de SEQ ID N° 1 du nucléotide 2322 au nucléotide 3404) représentées à la figure 2, eryBIV (séquence de SEQ ID N° 6 du nucléotide 242 au nucléotide 1207), eryBV (séquence de SEQ ID N° 6 du nucléotide 1210 au nucléotide 2454), eryCVI (séquence de SEQ ID N° 6 du nucléotide 2510 au nucléotide 3220), eryBVI (séquence de SEQ ID N° 6 du nucléotide 3308 au nucléotide 4837), eryCIV (séquence de SEQ ID N° 6 du nucléotide 4837 au nucléotide 6039), eryCV (séquence de SEQ ID N° 6 du nucléotide 6080 au nucléotide 7546) ou eryBVII (séquence de SEQ ID N° 6 du nucléotide 7578 au nucléotide 8156) représentées à la figure 3, pour synthétiser des métabolites secondaires hybrides chez *Sac. erythraea*.

Par métabolites secondaires hybrides, on entend soit des analogues de l'érythromycine, c'est-à-dire des dérivés de l'érythromycine ayant une ou plusieurs modifications portant sur la partie sucre et possédant une activité antibiotique, soit des précurseurs de l'érythromycine tels que le 6-désoxyérythronolide B ou l'érythronolide B auxquels sont attachés un ou plusieurs résidus sucre modifiés ou non et ne possédant pas d'activité antibiotique. Le résidu sucre modifié peut être par exemple, le 4-céto-L-mycarose.

La synthèse de métabolites secondaires hybrides chez *Sac. erythraea* par utilisation de séquences d'ADN eryB ou eryC de l'invention peut être réalisée par exemple, par l'inactivation d'un ou plusieurs gènes eryB ou eryC ci-dessus et l'introduction d'un ou plusieurs gènes exogènes ou de leurs dérivés obtenus par exemple par mutagénèse, ayant des séquences nucléotidiques codant pour des enzymes ayant la même fonction chez des souches productrices d'autres macrolides, par exemple la tylosine, la picromycine ou la méthymycine. En particulier, l'introduction de gènes exogènes peut être effectuée par intégration d'une séquence d'ADN obtenue selon la méthodologie du "DNA shuffling" (Stemmer, 1994) ou par la construction d'une séquence d'ADN chimère, par

exemple à partir d'une séquence eryB ou eryC de l'invention intervenant dans le transfert d'un résidu sucre, par exemple la séquence eryCIII ou eryBV, et de gènes homologues isolés à partir de souches productrices de macrolides, par exemple
5 *Streptomyces fradiae*, *Streptomyces olivaceus*, *Streptomyces venezuelae* ou *Streptomyces antibioticus*.

L'invention concerne aussi l'utilisation d'au moins l'une des séquences d'ADN choisie parmi les séquences eryBII (séquence complémentaire de SEQ ID N° 1 du nucléotide 48 au
10 nucléotide 1046), eryCIII (séquence complémentaire de SEQ ID N° 1 du nucléotide 1046 au nucléotide 2308) ou eryCII (séquence complémentaire de SEQ ID N° 1 du nucléotide 2322 au nucléotide 3404) représentées à la figure 2, eryBIV (séquence de SEQ ID N° 6 du nucléotide 242 au nucléotide 1207), eryBV
15 (séquence de SEQ ID N° 6 du nucléotide 1210 au nucléotide 2454), eryCVI (séquence de SEQ ID N° 6 du nucléotide 2510 au nucléotide 3220), eryBVI (séquence de SEQ ID N° 6 du nucléotide 3308 au nucléotide 4837), eryCIV (séquence de SEQ ID N° 6 du nucléotide 4837 au nucléotide 6039), eryCV
20 (séquence de SEQ ID N° 6 du nucléotide 6080 au nucléotide 7546) ou eryBVII (séquence de SEQ ID N° 6 du nucléotide 7578 au nucléotide 8156) représentée à la figure 3 ou d'un fragment de cette séquence, comme sondes d'hybridation.

Les séquences d'ADN eryB ou eryC de l'invention peuvent
25 être utilisées pour constituer des sondes d'hybridation d'au moins 19 nucléotides, permettant d'isoler des gènes homologues dans des souches productrices de macrolides en utilisant les méthodes classiques d'hybridation d'acides nucléiques immobilisées sur des filtres ou d'amplification
30 par PCR, selon les conditions décrites par Sambrook et al. (1989).

L'invention concerne particulièrement l'utilisation de la séquence d'ADN eryCIII représentée à la figure 2 (séquence complémentaire de SEQ ID N° 1 du nucléotide 1046 au nucléo-
35 tide 2308 = séquence complémentaire de SEQ ID N° 4) comme sonde d'hybridation pour isoler des gènes homologues responsables de la glycosylation de la macrolactone chez une souche productrice de macrolide.

L'invention concerne plus particulièrement l'utilisation ci-dessus, dans laquelle les gènes homologues sont les gènes de la biosynthèse de l'oléandomycine chez *S. antibioticus*.

L'invention décrit, à titre d'exemple, l'utilisation de
5 la séquence du gène *eryCIII* comme sonde d'hybridation pour isoler des gènes homologues dans une souche productrice d'oléandomycine. La sonde *eryCIII* utilisée a permis d'isoler les gènes *oleG1* et *oleG2* codant pour des glycosyltransférases chez *S. antibioticus* impliquées dans le transfert de la
10 désosamine et de l'oléandrose sur le noyau lactone.

La caractérisation fonctionnelle des gènes *oleG1* et *oleG2* a permis de définir l'organisation de la partie droite du cluster des gènes de la biosynthèse de l'oléandomycine chez *S. antibioticus*.

15 L'invention a donc pour objet une séquence d'ADN isolée représentée à la figure 22 (séquence de SEQ ID N° 15) correspondant à une région du cluster de gènes de la biosynthèse de l'oléandomycine comprenant :

- la séquence correspondant à l'ORF *oleP1* du nucléotide 184
20 au nucléotide 1386,
- la séquence correspondant à l'ORF *oleG1* du nucléotide 1437 au nucléotide 2714 codant pour une activité glycosyltransférase,
- la séquence correspondant à l'ORF *oleG2* du nucléotide 2722
25 au nucléotide 3999 codant pour une activité glycosyltransférase,
- la séquence correspondant à l'ORF *oleM* du nucléotide 3992 au nucléotide 4720 (= séquence de SEQ ID N° 20) et
- la séquence correspondant à l'ORF *oleY* du nucléotide 4810
30 au nucléotide 5967.

La séquence d'ADN ci-dessus montrée à la figure 22 (séquence de SEQ ID N° 15) est une séquence d'ADN génomique qui peut être obtenue par exemple à partir d'un cosmide couvrant la partie droite du cluster de gènes de la biosyn-
35 thèse de l'oléandomycine par hybridation avec une sonde *eryCIII*, selon les conditions opératoires dont une description détaillée est donnée plus loin.

L'invention a plus particulièrement pour objet une

séquence d'ADN isolée représentée à la figure 22 choisie parmi la séquence correspondant à l'ORF *oleG1* (séquence de SEQ ID N° 15 du nucléotide 1437 au nucléotide 2714 codant pour une activité glycosyltransférase et la séquence 5 correspondant à l'ORF *oleG2* (séquence de SEQ ID N° 15 du nucléotide 2722 au nucléotide 3999) codant pour une activité glycosyltransférase.

L'invention a tout particulièrement pour objet une séquence d'ADN isolée ci-dessus correspondant à l'ORF *oleG1* 10 (séquence de SEQ ID N° 15 du nucléotide 1437 au nucléotide 2714) codant pour une activité désosaminyltransférase, ainsi qu'une séquence d'ADN isolée ci-dessus correspondant à l'ORF *oleG2* (séquence de SEQ ID N° 15 du nucléotide 2722 au nucléotide 3999) codant pour une activité oléandrosyl- 15 transférase.

La séquence correspondant à l'ORF *oleG1* code pour un polypeptide ayant 426 acides aminés (séquence de SEQ ID N° 17) et la séquence correspondant à l'ORF *oleG2* code pour un polypeptide ayant 426 acides aminés (séquence de SEQ ID 20 N° 18).

La mise en évidence des activités enzymatiques respectives indiquées ci-dessus a été réalisée par altération du gène correspondant telle qu'illustrée plus loin dans la partie expérimentale.

25 L'invention a aussi pour objet le polypeptide codé par la séquence d'ADN correspondant à l'ORF *oleG1* et ayant une activité désosaminyltransférase (séquence de SEQ ID N° 17) et le polypeptide codé par la séquence d'ADN correspondant à l'ORF *oleG2* et ayant une activité oléandrosyltransférase 30 (séquence de SEQ ID N° 18).

Les polypeptides ci-dessus dénommés respectivement *OleG1* et *OleG2* peuvent être obtenus par les méthodes connues indiquées ci-dessus.

L'invention a aussi pour objet un procédé de préparation 35 de métabolites secondaires hybrides chez *Sac. erythraea* dans lequel :

- on isole une séquence ADN contenant au moins une séquence *eryB* ou une séquence *eryC* du cluster de gènes de la bio-

- synthèse de l'érythromycine représentée à la figure 2
(séquence complémentaire de SEQ ID N° 1) ou à la figure 3
(séquence de SEQ ID N°6),
- on crée une modification dans la dite séquence et on
- 5 obtient une séquence altérée,
- on intègre la séquence altérée dans le chromosome de la souche hôte et on obtient une souche modifiée,
 - on cultive la souche modifiée dans des conditions permettant la formation du métabolite secondaire hybride et
- 10 - on isole le métabolite secondaire hybride.

La modification de la séquence d'ADN peut être réalisée par exemple par une addition et/ou par une délétion de séquences d'ADN d'au moins un nucléotide, dans une séquence eryB ou eryC de l'invention qui code pour l'une des enzymes

15 correspondantes indiquées ci-dessus.

L'intégration de la séquence altérée dans la souche hôte peut être réalisée par exemple par la méthodologie de la recombinaison homologue qui peut être effectuée selon le schéma montré à la figure 4 et conduit à la génération de

20 mutants chromosomiques de souches *Sac. erythraea* que l'on cultive ensuite selon les méthodes générales connues de culture cellulaire.

L'invention a particulièrement pour objet le procédé ci-dessus dans lequel la séquence ADN code pour l'une des

25 enzymes choisie parmi une

- dTDP-4-céto-D-6-désoxyhexose 2,3-réductase,
 - désosaminyltransférase,
 - dTDP-4-céto-D-6-désoxyhexose 3,4-isomérase,
 - dTDP-4-céto-L-6-désoxyhexose 4-réductase,
- 30 - mycarosyltransférase,
- dTDP-D-6-désoxyhexose 3-N-méthyltransférase,
 - dTDP-4-céto-L-6-désoxyhexose 2,3-déshydratase,
 - dTDP-D-6-désoxyhexose 3,4-déshydratase,
 - dTDP-D-4,6-didésoxyhexose 3,4-réductase ou
- 35 - dTDP-4-céto-D-6-désoxyhexose 3,5 épimérase.

L'invention a plus particulièrement pour objet le procédé ci-dessus dans lequel l'altération de la séquence résulte dans l'inactivation d'au moins l'une des enzymes

indiquées ci-dessus.

L'inactivation d'au moins l'une des enzymes est mise en évidence, d'une part par l'absence de production d'érythromycine, d'autre part par l'accumulation de précur-
5 seurs de l'érythromycine tels que le 6-désoxyérythronolide B, l'érythronolide B ou le 3- α -mycarosyl érythronolide B et/ou l'accumulation de métabolites secondaires hybrides tels que définis précédemment dans les surnageants de cultures des souches modifiées correspondantes.

10 L'invention concerne tout particulièrement le procédé ci-dessus dans lequel l'enzyme inactivée est une dTDP-4-céto-L-6-désoxyhexose 4-réductase ou dans lequel l'enzyme inactivée est une dTDP-D-6-désoxyhexose 3,4-déshydratase ou dans lequel l'enzyme inactivée est une mycarosyltransférase
15 ou dans lequel l'enzyme inactivée est une dTDP-4-céto-L-6-désoxyhexose 2,3-réductase.

L'invention concerne aussi le procédé ci-dessus dans lequel le métabolite secondaire hybride isolé est un analogue de l'érythromycine choisi parmi la 4"-céto-érythromycine, la
20 4'-hydroxy-érythromycine ou la 3"-C désméthyl-2",3"-ène-érythromycine ou dans lequel le métabolite secondaire hybride isolé est le désosaminyl érythronolide B.

Des exemples de mise en oeuvre du procédé de l'invention sont donnés dans la partie expérimentale. L'accumulation de
25 métabolites secondaires hybrides dans des souches de *Sac. erythraea* modifiées est également décrite plus loin.

L'invention concerne aussi une souche de *Sac. erythraea* modifiée dans laquelle au moins l'une des enzymes choisie parmi une

- 30 - dTDP-4-céto-L-6-désoxyhexose 2,3-réductase,
- désosaminyltransférase,
- dTDP-4-céto-D-6-désoxyhexose 3,4-isomérase,
- dTDP-4-céto-L-6-désoxyhexose 4-réductase,
- mycarosyltransférase,
- 35 - dTDP-D-6-désoxyhexose 3-N-méthyltransférase,
- dTDP-4-céto-L-6-désoxyhexose 2,3-déshydratase,
- dTDP-D-6-désoxyhexose 3,4-déshydratase,
- dTDP-D-4,6-didésoxyhexose 3,4-réductase ou

- dTDP-4-céto-D-6-désoxyhexose 3,5 épimérase
est inactivée et produisant au moins un métabolite secondaire
hybride.

L'invention concerne particulièrement la souche de *Sac.*
5 *erythraea* modifiée BII92 dans laquelle une dTDP-4-céto-L-6-
désoxyhexose 2,3-réductase est inactivée et produisant la
3"-C désméthyl-2",3"-ène-érythromycine C, la souche de *Sac.*
erythraea modifiée BIV87 dans laquelle une dTDP-4-céto-L-6-
désoxyhexose 4-réductase est inactivée et produisant la 4"-
10 céto-érythromycine, la souche de *Sac. erythraea* modifiée
CIV89 dans laquelle une dTDP-D-6-désoxyhexose 3,4-déshydra-
tase est inactivée et produisant la 4'-hydroxyérythromycine D
ainsi que la souche de *Sac. erythraea* modifiée BV88 dans
laquelle une mycarosyltransférase est inactivée et produisant
15 du désoaminyl érythronolide B. Des constructions détaillées
des souches ci-dessus sont données plus loin dans la partie
expérimentale.

L'invention concerne aussi un procédé de préparation de
précurseurs de l'oléandomycine chez *S. antibioticus* dans
20 lequel
- on crée une altération de la séquence du gène choisie parmi
la séquence d'ADN correspondant à l'ORF *oleG1* (séquence de
SEQ ID N° 15 du nucléotide 1437 au nucléotide 2714)
et la séquence d'ADN correspondant à l'ORF *oleG2* (séquence
25 de SEQ ID N° 15 du nucléotide 2722 au nucléotide 3999) dans
le chromosome d'une souche hôte et obtient une souche
modifiée,
- on cultive la souche modifiée dans des conditions
permettant l'accumulation des précurseurs de l'oléandomycine
30 et
- on isole ces précurseurs.

L'altération de la séquence d'ADN peut être réalisée par
exemple par interruption du gène cible dans la souche
S. antibioticus, par exemple par intégration d'un plasmide
35 par la méthodologie de la recombinaison homologue et conduit
à la génération de mutants chromosomiques de la souche
sauvage.

L'invention concerne particulièrement un procédé ci-

dessus dans lequel l'altération est créée dans la séquence d'ADN correspondant à l'ORF *oleG1* (séquence de SEQ ID N° 15 du nucléotide 1437 au nucléotide 2714) et dont il résulte au moins l'élimination de l'activité désaminyltransférase et
5 l'accumulation du précurseur de l'oléandomycine 8,8a-désoxyoléandolide.

L'accumulation d'un précurseur non glycosylé de l'oléandomycine 8,8a-désoxyoléandolide observée par interruption du gène *oleG1* est due à un effet polaire transcription-
10 nel inactivant le gène *oleG2*.

Un exemple de mise en oeuvre du procédé ci-dessus est donné plus loin dans la partie expérimentale.

Matériels et méthodes générales.

15 1. Souches bactériennes, plasmides et conditions de croissance.

La souche *Sac. erythraea* utilisée pour la réalisation de l'invention est un variant phénotypique spontané dit "red variant" (Hessler et al., 1997) de la souche sauvage
20 *Sac. erythraea* NRRL 2338 dont la croissance est effectuée en routine soit sur milieu solide R2T2 (milieu R2T décrit par Weber et al., 1985 sans peptone), R2T20 (Yamamoto et al., 1986) ou M1-102 sur agar (Kaneda et al., 1962), soit en milieu liquide TSB (Oxoid) à 30°C.

25 La souche *Streptomyces lividans* 1326 (John Innes Culture Collection) décrite par Hopwood et al. (1983), utilisée pour la préparation de plasmides dépourvus d'origine de répllication d'*Escherichia coli* tels que pIJ702 et pIJ486, a été maintenue sur milieu solide R2YE(R5) (Hopwood et al., 1985).

30 La croissance des souches *E. coli* XL1-blue (Stratagene), JM110 (Stratagene) et DH5 α .MCR (GibcoBRL), utilisées pour les préparations de plasmides, a été effectuée en routine en milieu liquide 2 x YT ou LB ou en milieu solide LB sur agar, tels que décrits par Sambrook et al. (1989). La souche

35 *E. coli* XL1-blue est utilisée pour les clonages en routine. La souche JM110 est utilisée pour des clonages où l'on utilise des sites de restriction tels que *BclI*. La souche DH5 α .MCR est utilisée pour la préparation de plasmides

destinés à être introduits chez *Sac. erythraea* pour une transformation optimale.

La sélection des plasmides dans *E. coli* a été effectuée sur ampicilline (Sigma) à 100 µg/ml.

- 5 Les souches *Bacillus subtilis* ATCC 6633 ou *Bacillus pumilus* ATCC 14884 ont été utilisées comme souches indicatrices pour évaluer la production d'érythromycine dans des essais biologiques par antibiogramme.

Les plasmides Litmus28, pUC18 et pUC19 (New England
10 Biolabs) ont été utilisés en routine pour les sous-clonages. Le vecteur pIJ702 (Katz et al., 1983) a été obtenu du John Innes Institute. Le vecteur pIJ486 (Ward et al., 1986) a été obtenu de C.J. Thompson (Université de Bâle, Suisse). Le phagmide pTZ18R a été obtenu de Pharmacia Biotech. Le vecteur
15 navette coli-streptomyces pUWL218 (Wehmeier, 1995) utilisé pour l'intégration chromosomique dans *Sac. erythraea* a été obtenu de W. Piepersberg (Université de Wuppertal, Allemagne).

2. Manipulation de l'ADN et séquençage.

Les méthodes générales de biologie moléculaire utilisées
20 sont décrites par Sambrook et al., 1989.

Les réactifs d'origine commerciale ont été utilisés incluant les enzymes de restriction (New England Biolabs et Boehringer Mannheim), le fragment de Klenow de l'ADN
polymerase I (Boehringer Mannheim). La trousse "DNA ligation
25 system" (Amersham) a été utilisée pour effectuer les ligations et la trousse Plasmid Midi kit (Quiagen) ou RPM kit (Bio101 Inc.) pour purifier l'ADN plasmidique.

La préparation de l'ADN du bactériophage λ a été réalisée selon Ausubel et al. (1995) et l'isolement de l'ADN
30 chromosomique de *Sac. erythraea* selon Hopwood et al. (1985).

La transformation de *S. lividans* et l'isolement des plasmides ont été effectués selon Hopwood et al. (1985).

3. Préparation de l'érythronolide B et du 3- α -mycarosyl érythronolide B.

- 35 L'érythronolide B et le 3- α -mycarosyl érythronolide B ont été purifiés à partir d'extraits de culture du mutant eryCI (clone WHB2221 décrit par Dhillon et al., 1989) par chromatographie sur gel d'aminopropyl (LichroprepNH2 25-40 µ,

Merck) avec un gradient d'élution par des mélanges chlorure de butyle/chlorure de méthylène successifs (100:0, 80:20, 50:50 et 20:80) suivi d'un gradient d'élution linéaire par le mélange chlorure de butyle/méthanol variant de 99:1 à 90:10.

- 5 Les fractions contenant les produits attendus sont amenées à sec sous vide puis analysées par chromatographie en couche mince (ccm). L'érythronolide B est ensuite cristallisé dans un mélange acétate d'éthyle/hexane puis recristallisé dans l'éthanol. Le 3- α -mycarosyl érythronolide B est cristallisé
10 deux fois dans un mélange acétate d'éthyle/hexane.

Milieux cités.

1. R2T2 :

Pour 1 litre de solution aqueuse : sucrose 103 g ; K_2SO_4 0,25g ; extrait de levure 6,5 g ; tryptone 5,0 g ; bactoagar
15 22,0 g ; eau distillée qsp 860 ml. La solution est stérilisée par autoclavage pendant 30 minutes à 120°C. Au moment de l'utilisation, les solutions stériles suivantes sont ajoutées : 20 ml de glucose à 50 % ; 25 ml de Tris-HCl 1M, pH7,0 ; 5 ml de KH_2PO_4 à 0,5 % ; 2,5 ml de NaOH 1N ; 50 ml
20 de $CaCl_2$ 1M ; 50 ml de $MgCl_2 \cdot 6H_2O$ 1M et 2 ml de solution de "trace elements" (Hopwood et al., 1985).

2. R2T20 :

Pour un litre de solution aqueuse : milieu R2T2 contenant 206 g de sucrose.

25 3. M1-102 (Kaneda et al., 1962) :

Pour 1 litre de solution aqueuse : glucose 5 g ; sucre brun commercial 10 g ; tryptone 5 g ; extrait de levure 2,5 g ; Versène 36 mg ; eau courante 1000 ml ; pH final ajusté à 7,0 à 7,2 avec KOH. La solution est stérilisée par autoclavage
30 pendant 30 minutes à 120°C.

4. R2YE(R5) (Hopwood et al., 1985) :

Pour 1 litre de solution aqueuse : sucrose 103 g ; K_2SO_4 0,25 g ; $MgCl_2 \cdot 6H_2O$ 10,12 g ; casaminoacides 0,1 g ; solution de "trace elements" 2 ml ; extrait de levure 5 g ; TES
35 5,72 g ; bactoagar 15 g ; eau distillée qsp 940 ml. La solution est stérilisée par autoclavage pendant 30 minutes à 120°C. Au moment de l'utilisation, les solutions stériles suivantes sont ajoutées : 10 ml de KH_2PO_4 0,5 % ; 20 ml de

CaCl_2 1M ; 15 ml de L-proline à 20 % ; 20 ml de glucose à 50 % et 1 ml de CuCl_2 10mM.

5. 2 x TY :

Pour 1 litre de solution aqueuse : tryptone 10 g ; extrait de 5 levure 10 g ; NaCl 5 g.

6. Tampon PT :

Pour 1 litre de solution aqueuse : sucrose 100 g ; K_2SO_4 0,25 g ; $\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$ 5,1 g ; solution de "trace elements" 2 ml ; eau distillée qsp 875 ml. La solution est stérilisée 10 par autoclavage pendant 30 minutes à 120°C. Au moment de l'utilisation, les solutions stériles suivantes sont ajoutées : 5 ml de CaCl_2 et 20 ml de TES 5,3 %.

7. Sucrose-succinate (Caffrey et al., 1992) :

Pour 1 litre de solution aqueuse : sucrose 0,2 M ; acide 15 succinique 20 mM ; phosphate de potassium 20 mM (pH 6,6) ; sulfate de magnésium 5 mM ; nitrate de potassium 100 mM ; solution de "trace elements" 2 ml. La solution est stérilisée par autoclavage pendant 30 minutes à 120°C.

Les figures ci-annexées illustrent certains aspects de 20 l'invention.

La figure 1 représente la voie de biosynthèse de l'érythromycine A.

La figure 2 représente la séquence nucléotidique (séquence directe et complémentaire de SEQ ID N° 1) de la 25 région eryG-eryAIII du cluster de gènes de la biosynthèse de l'érythromycine comprenant les ORFs 7, 8 et 9 et leurs séquences protéiques déduites.

La figure 3 représente la séquence nucléotidique (séquence de SEQ ID N° 6) de la région eryAI-eryK du cluster 30 de gènes de la biosynthèse de l'érythromycine comprenant les ORFs 13 à 19 et leurs séquences protéiques déduites.

La figure 4 représente le schéma de substitution de gène par recombinaison homologue.

La figure 5A représente l'organisation de la partie 35 gauche du cluster des gènes de la biosynthèse de l'érythromycine chez *Sac. erythraea* dont les ORFs 1 à 9 sont indiquées par des flèches ainsi qu'une carte de restriction des plasmides pK62, pBCK1, pKB22, pBK44, pBIISB, pEco2 et pK23,

généérés à partir du clone génomique λ SE5.5. (Abréviations des enzymes de restriction : B, *Bam*HI ; Bc, *Bcl*I ; Bg, *Bgl*II ; E, *Eco*RI ; K, *Kpn*I ; M, *Mlu*I ; P, *Pst*I ; S, *Sac*I ; Sa, *Sal*I.)

La figure 5B représente l'organisation de la partie
5 droite du cluster des gènes de la biosynthèse de l'érythro-
mycine chez *Sac. erythraea* dont les ORFs 13 à 21 sont
indiquées par des flèches ainsi qu'une carte de restriction
des plasmides pBK6-12, pCN9, pNCO28, pNB49, pNCO62, pPSP4,
pNCO62X et pBAB18. (Abréviations des enzymes de restriction :
10 B, *Bam*HI ; Ba, *Bal*I ; Bc, *Bcl*I ; C, *Cla*I ; E, *Eco*RI ; K,
*Kpn*I ; N, *Nco*I ; Ns, *Nsi*I ; P, *Pst*I ; Pv, *Pvu*II ; S, *Sac*I ;
Sc, *Sca*I ; Sh, *Sph*I ; Sp, *Spe*I ; X, *Xba*I ; Xh, *Xho*I).

La figure 6A représente le schéma de construction du
plasmide pBIIA.

15 La figure 6B représente une carte de restriction du
plasmide pUWL218.

La figure 6C représente une carte de restriction du
plasmide pBIIA.

La figure 7A représente le schéma de construction du
20 plasmide pdel88.

La figure 7B représente le schéma de construction du
plasmide pdel88A.

La figure 7C représente le schéma de construction du
plasmide pOBB.

25 La figure 7D représente le schéma de construction et une
carte de restriction du plasmide pCIIIA.

La figure 8A représente le schéma de construction du
plasmide pCIIA.

La figure 8B représente une carte de restriction du
30 plasmide pORT1.

La figure 8C représente une carte de restriction du
plasmide pCIIA.

La figure 9A représente le schéma de construction du
plasmide pBIVA.

35 La figure 9B représente une carte de restriction du
plasmide pBIVA.

La figure 10A représente le schéma de construction du
plasmide pBVA.

La figure 10B représente une carte de restriction du plasmide pBVΔ.

La figure 11A représente le schéma de construction du plasmide pPSTI.

5 La figure 11B représente une carte de restriction du plasmide pPSTI.

La figure 12A représente le schéma de construction du plasmide pXhoI.

10 La figure 12B représente une carte de restriction du plasmide pXhoI.

La figure 13A représente le schéma de construction du plasmide pCIVΔ.

La figure 13B représente une carte de restriction du plasmide pCIVΔ.

15 La figure 14A représente le schéma de construction du plasmide pCVΔ.

La figure 14B représente une carte de restriction du plasmide pCVΔ.

La figure 15 représente l'analyse par Southern blot des
20 souches mutantes BII92, CIII68, CII62, BIV87, BV88, CIV89 et CV90, comparativement à la souche sauvage "red variant" notée Wt. Pour chaque mutant, l'enzyme de restriction utilisée est indiquée en-dessous de chaque blot et la taille des bandes détectées devant chaque blot est estimée par rapport aux
25 marqueurs de poids moléculaire λ-HindIII et λ-BstEII (non détectables par auto-radiographie).

La figure 16 représente l'analyse par PCR des souches mutantes BII91, CIII68, CII62, BIV87, BV88, CIV89 et CV90, comparativement à la souche sauvage "red variant" notée Wt et
30 aux plasmides pBIIΔ, pCIIIΔ, pCIIΔ, pBIVΔ, pBVΔ, pCIVΔ et pCVΔ utilisés respectivement pour obtenir le mutant par recombinaison homologue. Les tailles des bandes détectées par coloration au bromure d'éthyldium sont estimées par rapport aux marqueurs de poids moléculaire φX174-HaeIII ou λ-BstEII.

35 La figure 17 représente l'analyse par CCM des métabolites produits par les souches mutantes BII92, CIII68, CII62, BIV87, BV88, CIV89 et CV90, comparativement aux produits standards érythromycine A (Er A), érythronolide B (EB) et

3- α -mycarosyl érythronolide B (MEB).

La figure 18 représente l'analyse par SDS-PAGE de la purification de la protéine EryCIII successivement après extraction à l'urée 7M (ligne 2), chromatographie Q Sépharose 5 (ligne 3), chromatographie Superdex (ligne 4), chromatographie Q source (ligne 6) avec des marqueurs standard de poids moléculaire (lignes 1 et 5);

La figure 19 représente l'analyse par CMM du test d'activité biologique de la protéine EryCIII, par incubation avec d-TDP-D-désosamine (ligne 2) ou avec d-TDP-D-désosamine et 3- α -mycarosyl érythronolide B (MEB) (ligne 3) comparative-ment au contrôle MEB (ligne 1) et au contrôle érythromycine A (ligne 4). Les pointillés marquent les zones montrant une activité antibiotique par autobiogramme sur *B. pumilus*.

La figure 20 représente la localisation des six cosmides (cosAB35, cosAB76, cosAB87, cosAB67, cosAB63 et cosAB61) couvrant l'ensemble du cluster des gènes de la biosynthèse de l'oléandomycine. Les fragments de restriction BamHI (noté B) hybridant avec les sondes notées str M, D, E et les fragments BamHI (3,5 kb et 2,7 kb) hybridant avec la sonde eryCIII sont montrés.

La figure 21 représente l'organisation de la partie droite du cluster des gènes de la biosynthèse de l'oléandomycine chez *S. antibioticus* dont les différentes ORFs (notées oleP1, oleG1, oleG2, oleM, oleY, oleP et oleB) sont indiquées par des flèches ainsi qu'une carte de restriction du plasmide pCO35-S et l'insert du plasmide pCO3 généré à partir du pCO35-S. La double flèche indique l'insert correspondant à la séquence de la figure 22 (abréviations des enzymes de restriction : B, BamHI ; Bg, BglII ; K, KpnI ; S, SacI ; Sh, SphI ; l'étoile indique qu'il ne s'agit pas d'un site unique).

La figure 22 représente la séquence nucléotidique (séquence de SEQ ID N° 15) de la région couvrant les gènes oleP1, oleG1, oleG2, oleM et oleY de la biosynthèse de l'oléandomycine et leurs séquences protéiques déduites.

EXEMPLE 1 : clonage et séquençage de la région eryG-eryAIII du cluster de gènes de la biosynthèse de l'érythromycine.

Un fragment d'ADN génomique de *Sac. erythraea* NRRL 2338 ayant > 20 kb en aval du gène *ermE* couvrant notamment les ORFs 3 à 9 et correspondant au clone λ SE5.5 ainsi que la séquence nucléotidique d'un fragment de 4,5 kb correspondant 5 à la région du cluster *ery* comprise entre 3,7 kb et 8,0 kb à partir de l'extrémité 3' du gène *ermE* et comprenant les ORFs 3, 4, 5 et 6 ont été décrits par Haydock et al. (1991).

En tenant compte de la carte de restriction montrée par Haydock et al. (1991), des sous-clones ont été dérivés du 10 clone λ SE5.5 par sous-clonage de fragments de restriction dans pUC19. Les plasmides pKB22, pBK44, pBIISB et pEco2 ont été ainsi générés selon la figure 5A de la façon suivante :

A partir de l'ADN du clone λ SE5.5 digéré par l'enzyme de restriction *KpnI*, les plasmides pK62 et pK66 ont été 15 directement construits par sous-clonage du fragment *KpnI* de 5,8 kb dans pUC19, le plasmide pK66 correspondant au même fragment *KpnI* sous-cloné avec une orientation inversée de l'insert par rapport au vecteur. Le plasmide pKB22 contenant un insert de 2,9 kb a été ensuite dérivé du plasmide pK66 par 20 excision du fragment *BamHI-BglII* (2,9 kb) couvrant l'ORF8 ainsi qu'une partie des ORFs 7 et 9 par digestion avec les enzymes de restriction *BamHI* et *BglII*. De la même façon, le plasmide pKB44 contenant un insert de 2,9 kb a été obtenu à partir du plasmide pK62 par excision du fragment *BamHI-BglII* 25 (2,9 kb) couvrant le gène *eryG* correspondant aux ORFs 5 et 6 et le gène *eryF* correspondant à l'ORF4.

Le plasmide pBIISB a été dérivé du plasmide pBK44 par sous-clonage dans pUC19 du fragment *SalI* de 600 pb obtenu à partir du plasmide pBK44 digéré par l'enzyme de restriction 30 *SalI* (figure 5A).

A partir de l'ADN du clone λ SE5.5 digéré par l'enzyme de restriction *EcoRI*, le plasmide pEco2 a été directement construit par sous-clonage du fragment *EcoRI* (2,2 kb) dans pUC19.

35 Les sous-clones pKB22, pBK44, pBIISB et pEco2 ainsi obtenus ont été ensuite séquencés. L'analyse a été faite sur des échantillons d'ADN plasmidique, préalablement purifié sur une colonne de Quiagen 100 (Quiagen), sur le séquenceur

automatique ABI prism 377. Les réactions de séquençage ont été réalisées par la méthode de Sanger (1977) en utilisant les amorces M13 conventionnelles ou des amorces synthétiques et des didésoxynucléosides triphosphate fluorescents et la polymérase Taq FS (Perkin Elmer) en présence de 5 % de diméthylsulfoxyde, les amorces synthétiques utilisées ayant les séquences suivantes :

| | | | |
|----|---------|----------------------|-------------------|
| | C3R2 | TCCTCGATGGAGACCTGCC | (SEQ ID N° 22) |
| | B2R1 | GAGACCATGCCCAGGGAGT | (SEQ ID N° 23) |
| 10 | C3S2 | TCTGGGAGCCGCTCACCTT | (SEQ ID N° 24) |
| | C2R1 | GACGAGGCCGAAGAGGTGG | (SEQ ID N° 25) |
| | C2S | GCACACCGGAATGGATGCG | (SEQ ID N° 26) |
| | fullC3S | CCGTTCGAGCTCTGAGGTAA | (SEQ ID N° 27) |
| | fullC3R | GCCCGAGCCGCACGTGCGT | (SEQ ID N° 28) et |
| 15 | C4 | TGCACGCGCTGCTGCCGACC | (SEQ ID N° 29). |

L'assemblage des données de séquence a été réalisé avec le logiciel AutoassemblerTM package (Applied Biosystem). Les séquences ont été analysées en utilisant l'ensemble des logiciels GCG (Devereux 1984).

20 Les séquences nucléotidiques obtenues ont permis d'établir la séquence nucléotidique de 3412 bp de la figure 2 (séquence directe et complémentaire de SEQ ID N° 1) dans laquelle trois ORFs (7, 8 et 9) ont été identifiées respectivement du nucléotide 8957 au nucléotide 7959, du nucléotide 10219 au nucléotide 8957 et du nucléotide 11315 au nucléotide 10233 (numérotés dans la figure 2 à partir du site *Bam*HI situé à l'extrémité 5' du gène *ermE*) (respectivement séquence complémentaire de SEQ ID N° 1 du nucléotide 48 au nucléotide 1046, du nucléotide 1046 au nucléotide 2308 et du nucléotide 2322 au nucléotide 3404) et correspondant respectivement aux gènes *eryBII*, *eryCIII* et *eryCII* selon Liu et Thorson (1994) dont les caractérisations fonctionnelles n'avaient pas encore été identifiées. Les trois ORFs 7, 8 et 9 ont la même orientation, la lecture se faisant à partir de la région 3' du gène *eryAIII*.

Des échantillons de *E. coli* XL1-blue contenant la région codante des ORFs ci-dessus ont été déposés à la Collection Nationale de Cultures de Microorganismes (CNCM) INSTITUT

PASTEUR, 25, Rue du Docteur Roux 75724 PARIS CEDEX 15 FRANCE,
le 16 juillet 1997 :

- le plasmide pK62 comprenant la séquence codante pour l'ORF7, l'ORF8 et une partie de l'ORF9 sous le numéro I-1897,
- 5 - le plasmide pEco2 comprenant la séquence codante pour l'ORF9 et une partie de l'ORF8 sous le numéro I-1899.

EXEMPLE 2 : construction du plasmide pBIIA.

Un plasmide d'intégration, dénommé pBIIA et portant une délétion dans le gène *eryBII* codant pour l'ORF7, a été
10 construit selon le schéma de la figure 6A.

Le fragment *BclI*-*BamHI* de 598 pb a été délété dans le plasmide pK62 obtenu à l'exemple 1 par digestion avec les enzymes *BclI* et *BamHI*. Le plasmide pBCK1 résultant a été ensuite digéré avec les enzymes de restriction *MluI* et *BglII*
15 de façon à déléter un fragment ayant 853 pb à l'intérieur de l'ORF7 du nucléotide 8011 au nucléotide 8863 de la séquence de la figure 2. Après remplissage des extrémités à l'aide du fragment de Klenow de l'ADN polymérase I, le plasmide contenant la délétion, a été religaturé et transformé dans
20 *E. coli* XL1-blue. A partir du plasmide p19BIIA ainsi généré, le fragment *KpnI*-*HindIII* (4,3 kb) qui porte la délétion a été sous-cloné dans le plasmide pUWL218 (figure 6B). La présence de la délétion de 853 pb du nucléotide 8011 au nucléotide 8863 dans le plasmide pBIIA ainsi généré (figure 6C) a été
25 confirmée par séquençage.

Le plasmide pBIIA a ensuite été transféré dans la souche *E. coli* DH5αMRC, puis utilisé pour transformer *Sac. erythraea*.

EXEMPLE 3 : construction d'une souche *Sac. erythraea ery BIIA*
30 **(BII92).**

La construction d'une souche *Sac. erythraea* dans laquelle le gène *eryBII* porte une délétion interne telle que celle introduite dans le plasmide pBIIA préparé à l'exemple 2 et le processus d'intégration ont été réalisés de la façon
35 suivante :

La préparation des protoplastes a été réalisée selon la méthode décrite par Weber et Losick (1988), en utilisant du PEG 3350 (Sigma) au lieu de PEG 1000 et un tampon P modifié

(dénommé PT) contenant $\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$ 28 mM et sans $\text{PO}_4\text{H}_2\text{K}$ au lieu des tampons P, L ou T décrits, selon les conditions opératoires suivantes :

Les cellules (au moins 10^8 spores) de *Sac. erythraea* "red 5 variant" (dont un échantillon a été déposé à la Collection Nationale de Cultures de Microorganismes (CNCM) INSTITUT PASTEUR, 25, Rue du Docteur Roux 75724 PARIS CEDEX 15 FRANCE, le 16 juillet 1997 sous le numéro I-1902) ont été mises à pousser dans 50 ml de milieu TBS pendant 3 à 5 jours à 30°C, 10 puis lavées dans du sucrose à 10,3 %. Les cellules ont été remises en suspension dans 50 ml de tampon PT contenant 2 à 5 mg/ml de lysozyme (Sigma), puis incubées à 30°C pendant 1 à 2 heures en désagrégeant les amas de mycélium toutes les 15 minutes jusqu'à conversion d'au moins 50 % du mycélium en 15 protoplastes. Les protoplastes ont été lavés avec 50 ml de tampon PT, remis en suspension dans 12,5 à 25 ml du même tampon, congelés lentement puis stockés à -80°C par aliquots de 200 μl .

Pour la transformation, un aliquot a été décongelé et 20 50 μl ont été prélevés puis transférés dans un tube de 15 ml. Un à 10 μg d'ADN plasmidique pBII Δ , préparé à l'exemple 2 à partir de la souche *E. coli* DH5 α MRC ont été mis en solution dans 5 à 10 μl de tampon TE (Tris HCl 10 mM pH 7,5, EDTA 1 mM) puis déposés sur la paroi du tube incliné auquel a été 25 ensuite ajouté 0,5 ml d'une solution de PEG 3350 dans le tampon PT préparée extemporanément à partir d'une solution aqueuse à 50 % que l'on dilue au demi dans le tampon 2 x PT. Après dilution avec 3 à 5 ml de tampon PT puis centrifugation à 2500 rpm pendant 15 mn, le culot a été dissocié dans 0,5 ml 30 de tampon PT et la suspension de protoplastes transformés ainsi obtenue a été immédiatement répartie sur 2 ou 3 boîtes R2T2 très sèches (3 h sous une hotte à flux laminaire). Les boîtes ont été ensuite incubées à 32°C pendant 16 à 24 h jusqu'à apparition du voile de régénération des protoplastes. 35 A partir d'une solution stock de thiostrepton (Sigma) à 50 mg/ml dans le DMSO, une quantité appropriée a été diluée dans 0,5 à 1 ml d'eau puis étalée sur les boîtes de façon à obtenir une concentration finale de 20 μg de thiostrepton/ml

de gélose. Après absorption complète de l'antibiotique, les boîtes ont été incubées à 32°C pendant 3 à 4 jours, ce qui permet la visualisation des transformants. Les boîtes ont encore été incubées plusieurs jours jusqu'à développement
5 complet des spores.

La sélection des intégrants correspondant au premier événement de recombinaison (figure 4) a été réalisée par réplification des boîtes sporulées à l'aide de velours ou par étalement d'une suspension des spores sur des boîtes R2T2
10 contenant du thiostrepton puis incubation à 32°C, ce qui permet la croissance des clones d'intégrants potentiels.

Pour la sélection de clones ayant subi un deuxième événement de recombinaison (figure 4), 5 à 10 clones résistants au thiostrepton obtenus ci-dessus ont été mis en
15 culture dans 8 ml de milieu liquide TSB à 30°C pendant 3 à 4 jours. 50 à 100 µl ont été prélevés et remis en culture dans les mêmes conditions. Après 4 cycles successifs de dilution et culture destinés à favoriser la perte du marqueur de résistance au thiostrepton, des protoplastes ont été
20 préparées à partir des cellules comme indiqué ci-dessus, de façon à chasser le plasmide. Les protoplastes ont ensuite été étalés sur des boîtes R2T2 de façon à obtenir des colonies individualisées dont la sensibilité au thiostrepton a été déterminée par réplique sur des boîtes R2T2 contenant du
25 thiostrepton.

Selon la position du deuxième événement de recombinaison par rapport au site de délétion (figure 4), on peut attendre que le phénotype des colonies sensibles au thiostrepton soit du type sauvage ou du type muté porteur de la délétion.

30 Parmi les colonies sensibles au thiostrepton, la sélection des mutants ayant le phénotype *ery*⁻ a été réalisée par antibiogramme sur la souche *B. pumilus* ATCC 14884 sensible à l'érythromycine. La souche *B. pumilus* a été utilisée comme souche indicatrice pour évaluer la production
35 d'érythromycine dans des essais biologiques par antibiogramme. Les colonies ont été étalées à l'anse de platine sur des boîtes R2T2, puis incubées pendant 3 à 4 jours à 30°C. Des zones d'agar où le mutant a poussé à confluence ont

ensuite été prélevées à l'emporte pièce puis placées sur des boîtes A-Merck recouvertes d'une surcouche de 4 ml de 0,5 x A Merck (Antibiotic agar N°1 Merck) inoculée d'une suspension de spores de *B. pumilus*, puis incubées une nuit à 37°C.

5 La présence de la délétion attendue dans le chromosome du mutant (délétion de 853 pb du nucléotide 8011 au nucléotide 8863 de la figure 2) a ensuite été confirmée par analyse génomique par Southern blot ainsi que par PCR de la façon suivante :

10 Pour l'analyse par Southern blot, le transfert d'ADN génomique, préalablement digéré avec l'enzyme de restriction appropriée, sur des membranes GeenscreenPlus (Dupont NEN) a été réalisé dans NaOH 0,4 M selon Ausubel et al. (1995). Les hybridations ont été effectuées en utilisant comme sonde

15 l'oligonucléotide marqué à son extrémité 5' en utilisant du [γ^{32} P]ATP (Amersham) et la polynucléotide kinase (Boehringer Mannheim) selon Sambroock et al. (1989), ayant la séquence suivante :

B2-S TTGGCGAAGTCGACCAGGTC (SEQ ID N° 30)

20 correspondant à la région d'ADN du début du gène *eryG* située de la position 4118 à la position 4137 de la séquence déposée dans la base EMBL sous la référence X60379 et décrite par Haydock et al. (1991). Les hybridations ont été effectuées avec un tampon d'hybridation rapide (Amersham) et les

25 conditions de lavage suivantes : 2 x 5 mn, 2 x SSC, 20°C ; 30 mn, 2 x SSC, 65°C ; 30 mn, 0,1 x SSC, 20°C.

Par hybridation Southern sur l'ADN génomique isolé selon Hopwood et al. (1985) puis digéré par l'enzyme de restriction *KpnI*, une bande de 5,8 kb à partir de la souche sauvage "red

30 variant" et une bande de 4,9 kb à partir du mutant BII92 ont été détectées. Les résultats montrés à la figure 15 indiquent la présence chez le mutant d'une délétion d'environ 900 pb dans cette région du chromosome.

Pour l'analyse par PCR, un échantillon de 100 μ l d'une

35 culture de 3 jours en milieu TSB a été centrifugé. Le culot obtenu a été remis en suspension dans 10 μ l de milieu TSB, puis utilisé pour l'amplification dans l'appareil genAmp PCR system 9600 (Perkin Elmer Cetus). Après chauffage de

l'échantillon pendant 3 mn à 94°C, les conditions d'amplification suivantes ont été utilisées : 94°C, 1 mn ; 55°C, 1 mn ; 72°C, 3 mn ; 30 cycles ; polymérase Ampli Taq (Perkin Elmer) en présence de diméthylsulfoxyde 10 % (v/v) 5 suivis d'une élongation de 3 mn à 72°C . L'amplification a été effectuée en utilisant l'oligonucléotide B2S ci-dessus et l'oligonucléotide ayant la séquence suivante
B2-R GCCGCTCGGCACGGTGAAGTTCA (SEQ ID N° 31)
correspondant à la séquence du brin complémentaire de la 10 région d'ADN située de la position 8873 à la position 8892 de la séquence de la figure 2 à laquelle ont été ajoutés trois nucléotides à l'extrémité 5' et permettant d'encadrer par amplification PCR la région portant la délétion interne à l'ORF7.

15 L'analyse par amplification par PCR sur des cellules entières a permis de détecter une bande d'environ 1 kb dans la souche sauvage et une bande de 0,16 kb dans le mutant BII92 de façon identique au signal obtenu avec le plasmide pBIIA. Les résultats montrés à la figure 16 confirment que la 20 délétion d'environ 900 pb détectée par l'analyse Southern est identique à celle portée par le plasmide pBIIA (853 pb).

La souche recombinante ainsi obtenue, désignée BII92, a été ensuite cultivée pour identifier les métabolites produits par la souche.

25 **EXEMPLE 4 : fermentation de la souche BII92 et identification des métabolites secondaires produits.**

Des extraits de bouillon de culture de la souche ont été analysés par chromatographie en couche mince (ccm) avec l'érythromycine A, l'érythronolide B et le 3- α -mycarosyl 30 érythronolide B comme standards.

La souche BII92 a été cultivée en erlen de 50 ml dans les conditions permettant une production optimale d'érythromycine A et de ses dérivés qui consistent à effectuer une préculture cellulaire à 28°C pendant 48 heures 35 dans le milieu EP1 (Solulys L-Corn steep liquor (Roquette frères) 5 g/l ; farine de soja déshuilée (Cargill) 10 g/l ; CO₃Ca 2 g/l ; ClNa 5 g/l ; pH = 6,8 ; glucose qsp 15 g/l ajouté après autoclavage), puis une culture pendant 72 heures

après dilution à 7 % v/v avec le milieu EP2 (farine de soja déshuillée 10 g/l ; CO_3Ca 0,2 g/l ; $\text{Cl}_2\text{Co}-6\text{H}_2\text{O}$ 1 mg/l ; pH = 6,8-7,0 ; glucose qsp 20 g/l ajouté après autoclavage).

Le surnageant de culture a été ensuite extrait à pH 9-10 avec de l'acétate d'éthyle. Les phases organiques ont été séchées sur SO_4Mg , amenées à sec sous pression réduite puis analysées par ccm sur gel de silice 60 F254 (Merck) [dichlorométhane/méthanol (90:10, v/v) ou éther isopropylique/méthanol/ NH_4OH à 25 % (75:35:2, v/v)]. De façon alternative, l'analyse a été réalisée par ccm sur des plaques de gel de silice greffées de type NH_2 F254 (Merk) [chlorure de butyle/méthanol (90:10, v/v)].

La révélation chimique des plaques a été effectuée par pulvérisation d'une solution de p-anisaldéhyde-acide sulfurique 98 %-éthanol (1:1:9, v/v), suivie de chauffage pendant quelques minutes à 80°C. Les activités antibiotiques potentielles ont été analysées par bioautographie directe des plaques de ccm sur agarensemencé de *B. pumilus* ATCC 14884.

Les résultats obtenus par révélation chimique (figure 17) montrent que la souche BII92 accumule préférentiellement l'érythronolide B comme attendu d'un mutant *eryB*.

Des métabolites mineurs de faible mobilité manifestant une activité antibiotique ont également été détectés. Ces métabolites ont été extraits à l'acétate d'éthyle et identifiés par chromatographie liquide à haute performance en phase inverse (RP-HPLC) couplée à la spectrométrie de masse. La RP-HPLC a été effectuée sur colonne (250 x 4,6mm) de Kromasil C18 5 μ en utilisant comme phase mobile le mélange acétonitrile/méthanol/acétate d'ammonium 0,065 M pH 6,7 (350:150:500, v/v) sur un chromatographe Waters équipé d'un spectromètre de masse Finningan TSQ 7000.

A côté de traces d'érythromycine A, B, C et D, 4 métabolites mineurs dénommés M1 à M4 ont été détectés :
- M1 donne un pic parent à m/z 704 et des produits de fragmentation à m/z 576 et m/z 158. La présence de désaminyl-érythronolide A (m/z 576) indique que la différence de m/z de 30 comparée à l'érythromycine A (m/z 734) ou de 16 comparée à l'érythromycine C (m/z 720) est portée par le résidu sucre

neutre. La structure proposée pour M1 est la 3"-C désméthyl-2",3"-ène-érythromycine C.

- M2 donne un pic parent à m/z 706 et des produits de fragmentation à m/z 576 et m/z 158. La présence de désaminyl-érythronolide A (m/z 576) indique que la différence de m/z de 28 comparée à l'érythromycine A (m/z 734) ou de 14 comparée à l'érythromycine C (m/z 720) est portée par le résidu sucre neutre. La structure proposée pour M2 est la 3"-C désméthyl-érythromycine C.
- 10 - M3 donne un pic parent à m/z 690 et des produits de fragmentation à m/z 560 et m/z 158. La présence de désaminyl-érythronolide B (m/z 560) indique que la différence de m/z de 28 comparée à l'érythromycine B (m/z 718) ou de 14 comparée à l'érythromycine D (m/z 704) est portée par le résidu sucre neutre. La structure proposée pour M3 est la 3"-C désméthyl-érythromycine D.
- M4 donne un pic parent à m/z 720 et des produits de fragmentation à m/z 576 et m/z 158. Le profil est identique à celui de l'érythromycine C (m/z 720) avec la présence de désosaminylérythronolide A (m/z 576) et la perte du résidu sucre aminé (m/z 158), mais le métabolite M4 n'a pas le même temps de rétention en RP-HPLC que l'érythromycine. La structure proposée pour M4 est la 3"-C désméthyl-érythromycine A.

- 25 La détection par SM-SM du métabolite mineur M1 ayant un sucre neutre insaturé (3"-C désméthyl-2",3"-ène-érythromycine C) indique que le gène eryBII code pour la dTDP-4-céto-L-6-désoxyhexose 2,3-réductase dans la voie de biosynthèse du dTDP-mycarose.
- 30 La souche BII92 a été déposée à la Collection Nationale de Cultures de Microorganismes (CNCM) INSTITUT PASTEUR, 25, Rue du Docteur Roux 75724 PARIS CEDEX 15 FRANCE, le 16 juillet 1997 sous le numéro I-1903.

EXEMPLE 5 : construction du plasmide pCIIIA.

- 35 L'ORF8 pouvant être traductionnellement couplée à l'ORF7 située en aval, une délétion en phase a été introduite de façon à éviter un effet polaire. Un plasmide d'intégration, dénommé pCIIIA porteur d'une telle délétion, a été construit

selon le schéma de la figure 7(A-D).

Une délétion *SalI* de 663 pb a été introduite dans l'ORF8 du nucléotide 9384 au nucléotide 10046 de la séquence de la figure 2 en sous-clonant dans le plasmide pUC19 les deux
5 fragments *SalI* (a : 794 pb et b : 631 pb montrés à la figure 5A) isolés à partir du plasmide pBK44 obtenu à l'exemple 1 pour générer le plasmide pdel88 (figure 7A). La présence de la délétion de 663 pb a été confirmée par séquençage. Le plasmide pdel88 a été ensuite soumis à deux sous-clonages
10 additionnels de façon à élargir les régions chromosomiques utilisables pour la recombinaison homologue des deux cotés du site de délétion. Le fragment *SacI* (450 pb) du plasmide pdel88 a d'abord été remplacé par le fragment *SacI* (1,1 kb) du plasmide pEco2 obtenu à l'exemple 1 pour générer le
15 plasmide pdel88A (figure 7B). Puis le fragment *EcoRI* (1,5 kb) portant la délétion dans l'ORF8 a été isolé du plasmide pdel88A et utilisé pour remplacer le fragment *EcoRI* (1,66 kb) porteur de l'ORF intacte dans le plasmide pOBB. Le plasmide pOBB, représenté à la figure 7C, correspond au plasmide pBK44
20 préparé à l'exemple 1 dans le site *PstI* duquel a été sous-cloné le fragment *PstI* de 4 kb du plasmide pIJ486 obtenu par digestion par l'enzyme de restriction *PstI* et porteur de l'origine de répllication streptomyces ainsi que du gène de résistance au thiostrepton. Le plasmide résultant pCIIIA
25 porte des régions chromosomiques pour la recombinaison homologue de 1,27 kb et 1,38 kb respectivement en amont et en aval du site de délétion. Le plasmide pCIIIA ainsi obtenu (figure 7D) a ensuite été transféré dans la souche *E. coli* DH5 α MRC, puis utilisé pour transformer *Sac. erythraea*.
30 **EXEMPLE 6** : construction d'une souche *Sac. erythraea eryCIIIA* (CIIIA68).

Une souche dans laquelle le gène *eryCIII* porte une délétion interne telle que celle introduite dans le plasmide pCIIIA obtenu à l'exemple 5 a été préparée par transformation
35 des protoplastes de *Sac. erythraea* avec le plasmide pCIIIA.

La préparation des protoplastes, le processus d'intégration et la sélection des mutants ayant le phénotype *ery⁻* ont été réalisés comme à l'exemple 3.

De plus, la présence de la délétion attendue dans le chromosome (délétion de 663 pb du nucléotide 9384 au nucléotide 10046 de la séquence de la figure 2) a été confirmée par analyse génomique par Southern blot ainsi que
5 par amplification par PCR selon les conditions décrites à l'exemple 3.

Par hybridation Southern sur l'ADN génomique digéré par l'enzyme de restriction *EcoRI*, en utilisant comme sonde l'oligonucléotide ayant la séquence suivante

10 C3-S ATGCGCGTCGTCTTCTCCTCCATG (SEQ ID N° 32)
correspondant au brin complémentaire de la région d'ADN située de la position 10196 à la position 10219 de la séquence de la figure 2, une bande de 2,2 kb à partir de la souche sauvage et une bande de 1,5 kb à partir du mutant
15 CIII68 ont été détectées. Les résultats montrés à la figure 15 indiquent la présence chez le mutant d'une délétion d'environ 700 pb dans cette région du chromosome.

L'amplification par PCR a été effectuée en utilisant l'oligonucléotide C3-S ci-dessus et l'oligonucléotide ayant
20 la séquence suivante

C3-R TCATCGTGGTTCTCTCCTTCC (SEQ ID N° 33)
correspondant à la séquence située de la position 8954 à la position 8974 de la séquence de la figure 2 permettant d'encadrer par amplification PCR la région portant la
25 délétion interne à l'ORF8. L'analyse par amplification par PCR a permis de détecter une bande d'environ 1,2 kb dans la souche sauvage et une bande d'environ 0,6 kb dans le mutant CIII68 de façon identique au signal obtenu avec pCIIIΔ. Les résultats montrés à la figure 16 confirment que la délétion
30 d'environ 700 pb détectée par l'analyse de Southern est identique à celle portée par le plasmide pCIIIΔ (663 pb).

La souche recombinante ainsi obtenue, désignée CIII68, a été ensuite cultivée pour identifier les métabolites produits par la souche.

35 **EXEMPLE 7 : fermentation de la souche CIII68 et identification des métabolites secondaires produits.**

La culture de la souche CIII68 et les analyses par ccm suivie de bioautographie ont été réalisées selon les

conditions indiquées à l'exemple 4.

Les résultats de ccm (figure 17) montrent que la souche CII68 accumule préférentiellement du 3- α -mycarosyl érythronolide B ainsi que de petites quantités d'érythronolide B
5 comme attendu d'un mutant eryC.

La séquence eryCIII présente une forte homologie avec d'autres glycosyltransférases putatives telles que DauH (43 % d'identité au niveau protéique) et DnrS (47 % d'identité) impliquées dans la biosynthèse de la daunorubicine chez
10 *S. peucetius* (Otten et al., 1995) et chez *Streptomyces* sp C5 (Dickens et al., 1996) ainsi que TylM2 (50 % d'identité) impliquée dans le transfert du mycaminose sur la tylactone dans la voie de biosynthèse de la tylosine chez *S. fradiae* (Gandecha et al., 1997).

15 Ces observations indiquent que gène eryCIII code pour la désosaminyltransférase dans la voie de biosynthèse de l'érythromycine.

EXEMPLE 8 : construction du plasmide pCIIA .

Un plasmide d'intégration, dénommé pCIIA et portant une
20 délétion dans le gène eryBII codant pour l'ORF9, a été construit selon le schéma de la figure 8A.

Le plasmide pK23 (figure 5A) a été obtenu par sous-clonage dans pUC19 du fragment *KpnI* de 10 kb isolé à partir de l'ADN du clone λ SE5.5 digéré par l'enzyme de restriction
25 *KpnI*.

Dans un premier temps, le vecteur navette pORT1, montré à la figure 8B, a été obtenu par sous-clonage du fragment *PstI* de 4kb isolé par digestion du plasmide pIJ486 avec l'enzyme de restriction *PstI* incluant le gène de résistance
30 au thiostrepton et le réplicon *Streptomyces*, dans le site *PstI* de pUC19.

Une délétion hors phase de 304 pb a été introduite dans l'ORF9 du nucléotide 10881 au nucléotide 11184 de la séquence de la figure 2 en sous-clonant le fragment *SacI-KpnI* (1,1 kb)
35 du plasmide pK23 avec le fragment *EcoRI-KpnI* (1,7 kb) du plasmide pEco2 obtenu à l'exemple 1 dans le plasmide pORT1 ci-dessus préalablement digéré avec les enzymes de restriction *SacI* et *EcoRI*. Le plasmide d'intégration pCIIA ainsi

obtenu (figure 8C) a été ensuite transféré dans la souche *E. coli* DH5 α MRC, puis utilisé pour transformer *Sac. erythraea*.

EXEMPLE 9 : construction d'une souche *Sac. erythraea eryCIIA* (CII62).

- 5 Une souche dans laquelle le gène *eryCII* porte une délétion interne telle que celle introduite dans le plasmide pCIIA obtenu à l'exemple 8 a été préparée par transformation des protoplastes de *Sac. erythraea* avec le plasmide pCIIA.

La préparation des protoplastes, le processus d'inté-
10 gration et la sélection des mutants ayant le phénotype *ery*⁻ ont été réalisés comme à l'exemple 3.

De plus, la présence de la délétion attendue dans le chromosome (délétion de 304 bp du nucléotide 10881 au nucléotide 11184 de la séquence de la figure 2) a été
15 confirmée par analyse génomique par Southern blot ainsi que par PCR selon les conditions décrites à l'exemple 3.

Par hybridation Southern sur l'ADN génomique digéré par l'enzyme de restriction *EcoRI*, en utilisant comme sonde l'oligonucléotide C3-S ayant la séquence ci-dessus, une bande
20 de 2,2 kb à partir de la souche sauvage et une bande de 1,8 kb à partir du mutant CII62 ont été détectées. Les résultats montrés à la figure 15 indiquent chez le mutant la présence d'une délétion d'environ 400 pb dans cette région du chromosome.

- 25 L'amplification par PCR a été effectuée en utilisant l'oligonucléotide ayant la séquence suivante
C2-S GGAATTCATGACCACGACCGATC (SEQ ID N° 34)
correspondant au brin complémentaire de la région d'ADN de la fin du gène *eryAIII* située de la position 20258 à la position
30 20280 de la séquence déposée dans la base EMBL sous la référence X62569 et décrite par Bevitt et al., 1992 et l'oligonucléotide ayant la séquence suivante
C2-R CGCTCCAGGTGCAATGCCGGGTGCAGGC (SEQ ID N° 35)
correspondant à la séquence située de la position 10558 à la
35 position 10585 de la séquence de la figure 2 permettant d'encadrer par amplification PCR la région portant la délétion interne à l'ORF9. L'analyse par amplification par PCR a permis de détecter une bande d'environ 760 pb dans la

souche sauvage et une bande d'environ 460 pb dans le mutant CII62 de façon identique au signal obtenu avec pCIIA. Les résultats montrés à la figure 16 confirment que la délétion d'environ 400 pb détectée par l'analyse Southern est
5 identique à celle portée par le plasmide pCIIA (304 pb).

La souche recombinante ainsi obtenue, désignée CII62, a été ensuite cultivée pour identifier les métabolites produits par la souche.

**EXEMPLE 10 : fermentation de la souche CII62 et identifi-
10 cation des métabolites secondaires produits.**

La culture de la souche CII62 et les analyses par ccm suivie de bioautographie ont été réalisées selon les conditions indiquées à l'exemple 4.

Les résultats de ccm (figure 17) montrent que la souche
15 CII62 accumule préférentiellement du 3- α -mycarosyl érythronolide B ainsi que de petites quantités d'érythronolide B comme attendu d'un mutant eryC.

La séquence eryCII présente une forte homologie avec des gènes impliqués dans les voies de biosynthèse de la dauno-
20 samine (DnrQ, 38 % d'identité au niveau protéique, Otten et al., 1995) et du mycamino (protéine codée par l'ORF1*, 40 % d'identité au niveau protéique, Gandecha et al., 1997) qui ont également besoin de transférer un groupement céto en position 3 à partir d'un carbone adjacent.

25 Ces observations indiquent que le gène eryCII code pour la dTDP-4-céto-D-6-désoxyhexose 3,4-isomérase dans la voie de biosynthèse de la dTDP-désosamine.

**EXEMPLE 11 : clonage et séquençage de la région eryAI-eryK du
cluster de gènes de la biosynthèse de l'érythromycine.**

30 Des cosmides contenant la région eryAI-eryK du cluster de gènes ery tel que le cosmide Cos6B, ont été isolés par screening d'une banque d'ADN génomique de *Sac. erythraea* dans le vecteur cosmidique pWE15 (Stratagene) en utilisant comme sonde un fragment d'ADN de 13,2 kb comprenant la totalité du
35 gène eryAI et correspondant à la région d'ADN comprise entre le site NcoI situé à la position 44382 de la séquence de la figure 3 et le site NcoI situé à la position 392 de la séquence X62569 (Bevitt et al., 1992). La sonde a été

préparée de la façon suivante : Dans un premier temps, le fragment *NcoI* de 13,2 kb a été isolé à partir du plasmide pBK25 décrit par Bevitt et al., 1992 et sous-cloné dans le site *SmaI* de pUC18 après remplissage des extrémités *NcoI* avec
 5 le fragment de Klenow. A partir du plasmide pNCO12 ainsi généré, le fragment de 13,2 kb a été isolé par digestion avec l'enzyme de restriction *NcoI*.

Le cosmide cos6B ainsi obtenu a été digéré par l'enzyme de restriction *NcoI* et les fragments résultants de 2,8 kb et
 10 6,1 kb ont été clonés dans le site *NcoI* du vecteur Litmus28 générant respectivement les plasmides pNCO28 et pNCO62 montrés à la figure 5B.

Le plasmide pNCO28 a été séquencé par génération de sous-clones en utilisant l'exonucléase III selon le protocole
 15 du fournisseur de la trousse Erase-a-Base Kit (Promega) en digérant par les enzymes de restriction respectivement *SacI/XbaI* et *NsiI/BamHI* pour la direction inverse. La séquence a été complétée en utilisant comme amorces les oligonucléotides synthétiques ayant les séquences suivantes
 20 644 GATCACGCTCTTCGAGCGGCAG (SEQ ID N° 36)
 645 GAACTCGGTGGAGTCGATGTC (SEQ ID N° 37) et
 650 GTTGTCGATCAAGACCCGCAC (SEQ ID N° 38)

Pour le séquençage du plasmide pNCO62, des matrices ont été générées par sonication de l'ADN selon Bankier et al.
 25 (1987) en utilisant pUC18 comme vecteur. La séquence a été complétée en utilisant comme amorces les oligonucléotides synthétiques ayant les séquences suivantes :

646 CATCGTCAAGGAGTTCGACGGT (SEQ ID N° 39)
 647 TGCGCAGGTCCATGTTACCCACGTT (SEQ ID N° 40)
 30 648 GCTACGCCCTGGAGAGCCTG (SEQ ID N° 41)
 649 GTCGCGGTGGAGAGCACGAC (SEQ ID N° 42) et
 874 GCCAGCTCGGCGACGTCCATC (SEQ ID N° 43).

Les jonctions *NcoI* ont été séquencées en utilisant comme matrice l'ADN du cosmide cos6B obtenu ci-dessus dont les
 35 régions recouvrant les sites *NcoI* ont été séquencées en utilisant les amorces ayant les séquences 644 et 645 indiquées ci-dessus.

De plus, un fragment *ClaI-NcoI* de 0,9 kb, contenant le

début de la séquence du gène *eryAI* et la partie 5' de l'ORF13, a été cloné dans pUC18. Ce fragment a été préparé de la façon suivante : Le plasmide pBK6-12 représenté à la figure 5B a d'abord été généré par sous-clonage dans le
5 phagmide pTZ18R du fragment *KpnI* de 4,5 kb isolé à partir du plasmide pBK25 décrit par Bevitt et al., 1992. Le sous-clone pCN9 a ensuite été généré par sous-clonage du fragment *ClaI-NcoI* de 0,9 kb isolé à partir du plasmide pBK6-12 dans le site *SmaI* de pUC19, après remplissage des extrémités à l'aide
10 du fragment de Klenow. Le plasmide pCN9 ainsi obtenu (figure 5B) a été séquencé. Des matrices ont été générées par sonication de l'ADN selon Bankier et al. (1987) en utilisant pUC18 comme vecteur. La séquence a été complétée en utilisant comme amorce l'oligonucléotide ayant la séquence suivante :
15 875 CGACGAGGTCGTGCATCAG (SEQ ID N° 44).

Le séquençage de l'ADN est réalisé par la méthode de Sanger (1977) en utilisant un séquenceur automatisé sur les matrices d'ADN double brin avec le séquenceur Applied Biosystem 373 A. L'assemblage des données de séquence a été
20 réalisé avec le logiciel SAP (Staden, 1984). Les séquences ont été analysées en utilisant le logiciel GCG (Devereux, 1984).

Les séquences nucléotidiques obtenues ont permis d'établir la séquence nucléotidique de 8160 bp de la figure 3
25 (séquence de SEQ ID N° 6) dans laquelle sept ORFs (13-19) ont été identifiées respectivement du nucléotide 43841 au nucléotide 44806, du nucléotide 44809 au nucléotide 46053, du nucléotide 46109 au nucléotide 46819, du nucléotide 46907 au nucléotide 48436, du nucléotide 48436 au nucléotide 49638, du
30 nucléotide 49679 au nucléotide 51145 et du nucléotide 51177 au nucléotide 51755 (numérotés dans la figure 3 à partir du site *BamHI* situé à l'extrémité 5' du gène *ermE*) (respectivement séquence de SEQ ID N° 6 du nucléotide 242 au nucléotide 1207, du nucléotide 1210 au nucléotide 2454, du nucléotide
35 2510 au nucléotide 3220, du nucléotide 3308 au nucléotide 4837, du nucléotide 4837 au nucléotide 6039, du nucléotide 6080 au nucléotide 7546 et du nucléotide 7578 au nucléotide 8156) et correspondant respectivement aux gènes *eryBIV*,

eryBV, eryCVI, eryBVI, eryCIV, eryCV et eryBVII, selon Liu et Thorson (1994) dont les caractérisations fonctionnelles n'avaient pas encore été identifiées. Les sept ORFs (13-19) sont dans la même direction, la lecture se faisant à partir 5 de la région 5' du gène eryAI.

Des échantillons de *E. coli* XL1-blue contenant la région codante des ORFs ci-dessus ont été déposés à la Collection Nationale de Cultures de Microorganismes (CNCM) INSTITUT PASTEUR, 25, Rue du Docteur Roux 75724 PARIS CEDEX 15 FRANCE, 10 le 16 juillet 1997 :

- le plasmide pBK6-12 comprenant la séquence codante pour l'ORF13 et pour une partie de l'ORF14 sous le numéro I-1898
- le plasmide pNCO28 comprenant la séquence codante pour les ORFs 14 et 15 ainsi que pour une partie des ORFs 13 et 16 15 sous le numéro I-1901 et
- le plasmide pNCO62 comprenant la séquence codante pour les ORFs 17, 18 et 19 ainsi que pour une partie de l'ORF16 sous le numéro I-1900.

EXEMPLE 12 : construction du plasmide pBIVA.

20 L'ORF13 étant translationnellement couplé à l'ORF14 située en aval, une délétion en phase a dû être introduite. Un plasmide d'intégration, dénommé pBIVA et portant cette délétion, a été construit selon le schéma de la figure 9A.

Le plasmide pPSP4 (figure 5B) a d'abord été construit 25 par sous-clonage du fragment *PvuII-SpeI* (2,7 kb) isolé à partir du plasmide pBK6-12 obtenu à l'exemple 11 et du fragment *SpeI-PstI* (1,6 kb) isolé à partir du plasmide pNCO28 obtenu à l'exemple 11 dans le vecteur pUC19 préalablement digéré à l'aide des enzymes de restriction *SmaI* et *PstI*.

30 A partir du plasmide pPSP4, le plasmide p19BIVA a été généré en délétant le fragment *BclI-NcoI* de 510 pb interne à l'ORF13 et en lui substituant 45 pb venant d'un adaptateur synthétique de 54 pb. Cet adaptateur a été généré par appariement des 2 oligonucléotides complémentaires ayant les 35 séquences suivantes

SEQ A

AATTGATCAAGGTGAACACGGTCATGCGCAGGATCCTCGAGCGGAACTCCATGGGG

(SEQ ID N° 45) et

SEQ B

CCCCATGGAGTTCCGCTCGAGGATCCTGCGCATGACCGTGTTACCTTGATCAATT

(SEQ ID N° 46)

créant un site *Bcl*I et un site *Nco*I encadrant la séquence de
5 45 pb.

Pour l'appariement, les deux oligonucléotides ont été
mis à une concentration finale 1,8 μ M dans le tampon
d'hybridation NaCl 50 mM, Tris, HCl 20 mM pH 7,4, MgCl₂.6H₂O
2 mM, chauffés pendant 5 mn à 100°C puis refroidis lentement
10 à température ambiante. Après digestion avec les enzymes de
restriction *Nco*I et *Bcl*I, une ligation a été effectuée dans
le plasmide pPSP4 dont le fragment *Bcl*I-*Nco*I de 510 pb avait
préalablement été éliminé. A partir du plasmide p19BIVA ainsi
généré, le fragment *Sac*I-*Eco*RI (2,2 kb) portant l'ORF13
15 modifiée a été sous-cloné dans le plasmide pUWL218 préalable-
ment digéré avec les enzymes de restriction *Sac*I et *Eco*RI. Le
plasmide d'intégration pBIVA ainsi obtenu (figure 9B) a été
ensuite transféré dans la souche *E. coli* DH5 α MRC, puis
utilisé pour transformer *Sac. erythraea*.

20 **EXEMPLE 13 : construction d'une souche *Sac. erythraea* eryBIVA
(BIV87).**

Une souche dans laquelle le gène eryBIV porte une
délétion interne telle que celle introduite dans le plasmide
pBIVA obtenu à l'exemple 12 a été préparée par transformation
25 des protoplastes de *Sac. erythraea* avec le plasmide pBIVA.

La préparation des protoplastes, le processus
d'intégration et la sélection des mutants ayant le phénotype
ery⁻ ont été réalisés comme à l'exemple 3.

De plus, la présence de la délétion attendue dans le
30 chromosome (délétion de 510 bp du nucléotide 43872 au
nucléotide 44382 de la séquence de la figure 3) et son
remplacement par la séquence synthétique de 45 pb a été
confirmée par analyse génomique par Southern blot ainsi que
par PCR selon les conditions décrites à l'exemple 3.

35 Par hybridation Southern sur l'ADN génomique digéré par
l'enzyme de restriction *Xho*I, en utilisant comme sonde
l'oligonucléotide B4-R ayant la séquence suivante

B4-R AACTCGGTGGAGTCGATGTCGTCGCTGCGGAA (SEQ ID N° 47)

correspondant au brin complémentaire de la région d'ADN située de la position 44687 à la position 44718 de la séquence de la figure 3, une bande de 5,4 kb à partir de la souche sauvage et une bande de 2,7 kb à partir du mutant BIV87 ont été détectées. Les résultats montrés à la figure 15 indiquent la présence d'un site *XhoI* supplémentaire à une distance de 2,7 kb en amont du site *XhoI* situé à la position 47114 de la séquence de la figure 3 confirmant ainsi l'incorporation de l'adaptateur dans le chromosome de mutant, telle qu'attendue par l'incorporation de l'adaptateur synthétique ci-dessus utilisé pour générer le plasmide pBIVΔ.

L'amplification par PCR a été effectuée en utilisant l'oligonucléotide ayant la séquence suivante
B4-S CAATATAGGAAGGATCAAGAGGTTGAC (SEQ ID N° 48)
correspondant à la région d'ADN située de la position 43652 à la position 43678 de la séquence de la figure 3 et l'oligonucléotide B4-R ayant la séquence indiquée ci-dessus, permettant d'encadrer par amplification PCR la région portant la délétion interne à l'ORF13. L'analyse par amplification par PCR a permis de détecter une bande d'environ 1 kb dans la souche sauvage et une bande d'environ 500 pb dans le mutant BIV87 de façon identique au signal obtenu avec le plasmide pBIV87 (figure 16).

L'ensemble des résultats d'analyse par Southern et par PCR confirme la présence de la délétion de 510 pb et de l'adaptateur synthétique au niveau du chromosome du mutant.

La souche recombinante ainsi obtenue, désignée BIV87, a été ensuite cultivée pour identifier les métabolites produits par la souche.

30 EXEMPLE 14 : fermentation de la souche BIV87 et identification des métabolites secondaires produits.

La culture de la souche BIV87 et les analyses par ccm ont été réalisées selon les conditions indiquées à l'exemple 4.

35 Les résultats de ccm (figure 17) montrent que la souche BIV87 accumule préférentiellement de l'érythronolide B comme attendu d'un mutant *eryB*.

Des métabolites mineurs de faibles mobilités manifestant

une activité antibiotique ont également été détectés. Ces métabolites ont été extraits et analysés par RP-HPLC couplée à la spectrométrie de masse comme décrit à l'exemple 4.

Les résultats de spectre de masse indiquent que des
5 formes modifiées de l'érythromycine A, B, C et D ont été produites. Un métabolite majeur et 3 métabolites mineurs ont été détectés.

Le métabolite majeur M5 donne un pic parent à m/z 702 avec des produits de déshydratation et de fragmentation à m/z
10 684, m/z 560 et m/z 158 et correspond à l'élimination de 2 atomes d'hydrogène dans l'érythromycine D (m/z 704, m/z 686). La présence de désosaminyl érythronolide B (fragment m/z à 560) indique que la différence de masse est portée par le sucre neutre. La structure proposée pour ce métabolite est la
15 4"-céto érythromycine D.

Les métabolites mineurs donnent aussi un profil avec une différence de 2 dans les valeurs m/z respectivement :

- M6 (m/z à 718, m/z 700, m/z 576, m/z 158) au lieu de m/z 720, m/z 702 pour l'érythromycine C ;
- 20 - M7 (m/z à 732, m/z 714, m/z 576, m/z 158) au lieu de m/z 734, m/z 716 pour l'érythromycine A ;
- M8 (m/z à 716, m/z 698, m/z 560, m/z 158) au lieu de m/z 718, m/z 700 pour l'érythromycine B.

Les structures proposées sont respectivement la 4"-céto
25 érythromycine C pour M6, la 4"-céto érythromycine A pour M7 et la 4"-céto érythromycine B pour M8.

Ces observations indiquent que le gène eryBIV code pour la dTDP-4-céto-L-6-désoxyhexose 4-réductase dans la voie de biosynthèse du dTDP-mycarose.

30 La souche BIV87 a été déposée à la Collection Nationale de Cultures de Microorganismes (CNCM) INSTITUT PASTEUR, 25, Rue du Docteur Roux 75724 PARIS CEDEX 15 FRANCE, le 16 juillet 1997 sous le numéro I-1904

EXEMPLE 15 : construction du plasmide pBVΔ.

35 Un plasmide d'intégration, dénommé pBVΔ et portant une délétion dans le gène eryBV codant pour l'ORF14, a été construit selon le schéma de la figure 10A.

Une délétion de 726 pb a été générée dans l'ORF14 du

nucléotide 44963 au nucléotide 45688 de la séquence de la figure 3 par ligature du fragment *BclI*-*KpnI* (1,1 kb) isolé à partir du plasmide pBK6-12, obtenu à l'exemple 11, au fragment *KpnI*-*BamHI* (1,1 kb) isolé à partir du plasmide pNCO28, obtenu à l'exemple 11, dans le plasmide pUWL218 préalablement digéré par l'enzyme de restriction *BamHI*. Le plasmide d'intégration pBVA ainsi obtenu (figure 10B) a ensuite été transféré dans la souche *E. coli* DH5 α MRC, puis utilisé pour transformer *Sac. erythraea*.

10 **EXEMPLE 16 : construction d'une souche *Sac. erythraea* eryBVA (BV88).**

Une souche dans laquelle le gène eryBV porte une délétion interne telle que celle introduite dans le plasmide pBVA obtenu à l'exemple 15 a été préparée par transformation des protoplastes de *Sac. erythraea* avec le plasmide pBVA.

La préparation des protoplastes, le processus d'intégration et la sélection des mutants ayant le phénotype ery⁻ ont été réalisés comme à l'exemple 3.

De plus, la présence de la délétion attendue dans le chromosome (délétion de 726 pb du nucléotide 44963 au nucléotide 45688 de la séquence de la figure 3) a été confirmée par analyse génomique par Southern blot ainsi que par PCR selon les conditions décrites à l'exemple 3.

Par hybridation Southern sur l'ADN génomique digéré par l'enzyme de restriction *NcoI*, en utilisant comme sonde l'oligonucléotide ayant la séquence suivante :
B5-R TCCGGAGGTGTGCTGTCGGACGGACTTGTCCGGTCGGAAA (SEQ ID N° 49) correspondant au brin complémentaire de la région d'ADN située de la position 46060 à la position 46098 de la séquence de la figure 3, une bande de 2,7 kb à partir de la souche sauvage et une bande de 2,0 kb à partir du mutant BV88 ont été détectées. Les résultats montrés à la figure 15 indiquent chez le mutant la présence d'une délétion d'environ 700 pb dans cette région du chromosome.

35 L'amplification par PCR a été effectuée en utilisant l'oligonucléotide ayant la séquence suivante :
B5-S AGGAGCACTAGTCCGGGTACTGCTGACGTCCTT (SEQ ID N° 50) correspondant à la région d'ADN située de la position 44799 à

la position 44831 de la séquence de la figure 3 et l'oligonucléotide B5-R ayant la séquence indiquée ci-dessus, permettant d'encadrer par amplification PCR la région portant la délétion interne à l'ORF14. L'analyse par amplification
5 par PCR a permis de détecter une bande d'environ 1,3 kb dans la souche sauvage et une bande d'environ 570 pb dans le mutant BV88 de façon identique au signal obtenu avec le plasmide pBV88 (figure 16). Ces résultats confirment que la délétion de 710 pb détectée par analyse Southern est
10 identique à celle portée par le plasmide pBVA (726 pb).

La souche recombinante ainsi obtenue, désignée BV88, a été ensuite cultivée pour identifier les métabolites produits par la souche.

**EXEMPLE 17 : fermentation de la souche BV88 et identification
15 des métabolites secondaires produits.**

La culture de la souche BV88 et les analyses par ccm ont été réalisées selon les conditions indiquées à l'exemple 4.

Les résultats de ccm (figure 17) montrent que la souche BV88 accumule préférentiellement de l'érythronolide B comme
20 attendu d'un mutant eryB.

Des métabolites mineurs de faibles mobilités ont également été détectés puis extraits et identifiés par RP-HPLC couplée à la spectrométrie de masse selon les conditions utilisées à l'exemple 4.

25 Le spectre de masse montre la présence d'un métabolite ayant un pic parent à m/z 560 et des produits de déshydratation et de fragmentation à m/z 542 et m/z 158 pour lequel la structure proposée est le désosaminyl érythronolide B.

La séquence eryBV présente une forte homologie avec
30 d'autres glycosyltransférases ainsi qu'avec le gène eryCIII ci-dessus (60,7 % d'identité au niveau nucléotidique, 44 % au niveau protéique).

Ces observations indiquent que le gène eryBV code pour la mycarosyltransférase impliquée dans la biosynthèse de
35 l'érythromycine.

EXEMPLE 18 : construction d'un plasmide pCVIA (pPSTI).

Un plasmide d'intégration, dénommé pPSTI et portant une délétion dans le gène eryCVI codant pour l'ORF15, a été

construit selon le schéma de la figure 11A de la façon suivante :

Dans un premier temps, le plasmide pNB49 a été généré par traitement à l'exonucléase III du plasmide pNCO28 obtenu à l'exemple 11 préalablement digéré par les enzymes de restriction *NsiI* et *BamHI*. Le plasmide pNB49 (figure 5B) contenant les nucléotides 44382 à 46562 de la séquence de la figure 3, a été ensuite digéré à l'aide de l'enzyme de restriction *PstI* puis traité par la nucléase Mung Bean (NE Biolabs) comme décrit par Sambrook et al. (1989). Après religature et transformation dans *E. coli* XL1-Blue, les colonies résistantes à l'ampicilline ont été sélectionnées par analyse de restriction avec l'enzyme *PstI*. La perte du site *PstI* a été confirmée par séquençage d'un clone en utilisant l'amorce M13 inverse et la délétion du nucléotide 46364 de la séquence de la figure 3 a été observée créant un changement de phase dans l'ORF15 dans le plasmide pNB49Δ*Pst* ainsi généré. Le plasmide pIJ702 digéré avec l'enzyme de restriction *BglII* a été ensuite ligaturé au site *BglII* du plasmide pNB49Δ*Pst* générant le plasmide pPSTI. L'orientation de pIJ702 dans pPSTI a été confirmée par la présence d'un fragment d'ADN ayant 0,9 kb après digestion avec l'enzyme de restriction *SphI*. Le plasmide d'intégration pPSTI (figure 11B) ainsi obtenu a été transféré dans la souche *E. coli* DH5αMRC, puis utilisé pour transformer *Sac. erythraea*.

EXEMPLE 19 : construction d'une souche *Sac. erythraea eryCVIA* (Pst10).

Une souche dans laquelle le gène *eryCVI* porte une délétion interne telle que celle introduite dans le plasmide pPSTI obtenu à l'exemple 18 a été préparée par transformation des protoplastes de *Sac. erythraea* avec le plasmide pPSTI.

La préparation des protoplastes et le processus d'intégration ont été réalisés comme à l'exemple 3.

La sélection des mutants ayant le phénotype *ery*⁻ a été réalisée comme à l'exemple 3 en utilisant une souche *B. subtilis* sensible à l'érythromycine au lieu d'une souche *B. pumilus* comme souche indicatrice. La souche *B. subtilis* ATCC 6633 a été utilisée pour évaluer la production

d'érythromycine dans des essais biologiques sur des boîtes d'agar en milieu M1-102 inoculées avec le mutant à analyser et incubées pendant 3 jours à 30°C. Des zones d'agar recouvertes de bactéries ont ensuite été prélevées à
5 l'emporte pièce puis placées sur des boîtes 2 x TY recouvertes d'une surcouche de 5 ml d'agar en milieu TY contenant 200 µl d'une culture de *B. subtilis* ATCC 6633, puis incubées une nuit à 37°C.

L'absence de production d'érythromycine a été évaluée
10 également en présence de précurseurs ajoutés tels que l'érythronolide B ou le 3- α -mycarosyl érythronolide B par application de 10 µl d'une solution 10 mM de chaque métabolite sur les zones d'agar découpées suivie d'une incubation à 30°C pendant une nuit avant de recouvrir les boîtes de la
15 culture de *B. subtilis* comme indiqué ci-dessus. La souche *Sac. erythraea* sauvage "red variant" a été utilisée comme contrôle.

Après la transformation des protoplastes avec le plasmide pPSTI et la sélection des colonies résistantes au
20 thiostrepton, l'intégration dans le chromosome a été confirmée par analyse de Southern selon les méthodes générales décrites à l'exemple 3.

Un fragment d'ADN de 1269 pb correspondant à l'ORF14 généré par PCR en utilisant les oligonucléotides synthétiques
25 ayant les séquences suivantes :
14-1 GGGGGATCCCATATGCGGGTACTGCTGACGTCCTTCG (SEQ ID N° 51) et
14-2 GAAAAGATCTGCCGGCGTGGCGGCGGTGAGTTCCTC (SEQ ID N° 52)
a été utilisé comme sonde.

L'oligonucléotide 14-1 a été dessiné de façon à
30 introduire un site *Bam*HI et un site *Nde*I en amont de la séquence correspondant à la région d'ADN située de la position 44811 à la position 44833 de la séquence de la figure 3.

L'oligonucléotide 14-2 a été dessiné de façon à introduire un site *Bgl*III en aval de la séquence correspondant au
35 brin complémentaire de la région d'ADN située de la position 46027 à la position 46053 de la séquence de la figure 3.
L'ADN chromosomique préalablement digéré avec les enzymes de restriction *Cla*I et *Pst*I a montré les bandes attendues de

4 kb et 7 kb à partir de l'intégrant alors que la souche sauvage présentait la bande de 3 kb attendue.

Après cultures répétées des intégrants, les colonies individualisées obtenues ont été analysées pour la sensibilité au thiostrepton et la production d'érythromycine, puis l'intégration de la délétion attendue (délétion du nucléotide 46364 de la séquence de la figure 3) dans le chromosome d'un clone mutant ayant le phénotype *ery*⁻ (Pst10) a été confirmée par analyse de Southern. L'ADN chromosomique, isolé respectivement à partir de la souche sauvage et du mutant Pst10, a été digéré avec l'enzyme de restriction *Pst*I. L'hybridation avec la sonde *Pst*I-*Nco*I de 0,8kb (nucléotides 46368 à 47142 de la séquence de la figure 3) a donné le profil attendu avec une bande *Pst*I de 1kb correspondant aux nucléotides 46368 à 47397 de la séquence de la figure 3 à partir de la souche sauvage et avec une bande >20 kb à partir du mutant. La perte du site *Pst*I à la position 46368 ci-dessus a aussi été montrée après double digestion par les enzymes *Pst*I et *Nco*I, résultant en une bande *Pst*I-*Nco*I de 0,8 kb (nucléotide 46368 à 47142) à partir de la souche sauvage et une bande *Nco*I de 2,8 kb (nucléotide 44382 à 47142) avec le mutant.

La souche recombinante ainsi obtenue, désignée Pst10, a ensuite été cultivée pour identifier les métabolites produits.

25 EXEMPLE 20 : fermentation de la souche Pst10 et identification des métabolites secondaires produits.

La souche Pst10 a été cultivée dans le milieu sucrose-succinate décrit par Caffrey et al. (1992) pendant 3 jours à 30°C. Le surnageant de culture a été ensuite extrait à pH 9 avec de l'acétate d'éthyle. Les phases organiques obtenues ont été séchées sur SO_4Mg_2 puis amenées à sec sous pression réduite. Le résidu a été dissous dans le mélange acétonitrile-eau (1:1, v/v), puis a ensuite été analysé par spectrométrie de masse sur un spectromètre BioQ (Micromass, Manchester, UK) ou Finningan LCQ (Finningag, CA).

La production d'érythromycine A (*m/z* 734 et *m/z* 716) n'a pas été observée mais la présence d'érythronolide B (MK⁺ : *m/z* 441 et MNa⁺ : *m/z* 425) ainsi que de 3- α -mycarosyl

érythronolide B (MK⁺ : m/z 585 et MNa⁺ : m/z 569) mise en évidence caractérise la souche Pst10 comme un mutant eryC.

La séquence eryCVI présente une forte homologie avec d'autres méthyltransférases telles que SnoX impliquée dans la biosynthèse de la nogalamycine chez *S. nogalater* (numéro d'accèsion EMBL S52403) (55,5 % d'identité au niveau protéique), TylM1 impliquée dans la biosynthèse de la tylosine chez *S. fradiae* (numéro d'accèsion EMBL X81885) (65 % d'identité au niveau protéique) et SrmX impliquée dans la biosynthèse de la spiramycine chez *S. ambofaciens* (numéro d'accèsion EMBL S25204) (52,8 % d'identité au niveau protéique).

Ces observations indiquent que le gène eryCVI code pour la dTDP-D-6-désoxyhexose 3-N-méthyltransférase impliquée dans la voie de biosynthèse de la dTDP-D-désosamine.

EXEMPLE 21 : construction d'un plasmide pBVIA (pXhoI).

Un plasmide d'intégration, dénommé pXhoI et portant une délétion dans le gène eryBVI codant pour l'ORF16, a été construit selon le schéma de la figure 12A de la façon suivante :

Le fragment NcoI-XhoI (3,1 kb) du plasmide pNC062 obtenu à l'exemple 11 et contenant les nucléotides 47142 à 50254 de la séquence de la figure 3 a été sous-cloné dans les sites NcoI et XhoI du plasmide Litmus 28. Le plasmide pNC062X (figure 5B) ainsi généré a été digéré avec l'enzyme de restriction PstI puis traité avec l'ADN polymérase T4 (Boehringer Mannheim). Après religature et transformation dans *E. coli* XL1-Blue, la perte du site PstI au nucléotide 47397 de la séquence de la figure 3 a été confirmée par séquençage et une délétion de 60 pb du nucléotide 47337 au nucléotide 47397 de la séquence de la figure 3 ont été observés. Le plasmide pIJ702 digéré avec l'enzyme de restriction BglIII a été ensuite ligaturé au site BglIII de cette construction. L'orientation de pIJ702 dans la construction a été confirmée par la présence d'un fragment d'ADN ayant 4,3 kb après digestion avec l'enzyme de restriction XhoI. Le plasmide d'intégration pXhoI (figure 12B) ainsi obtenu a été utilisé pour transformer *Sac. erythraea*.

EXEMPLE 22 : construction d'une souche *Sac. erythraea eryBVIΔ* (Xho91).

Une souche dans laquelle le gène *eryBVI* porte une délétion interne telle que celle introduite dans le plasmide
5 pXhoI obtenu à l'exemple 21 a été préparée par transformation des protoplastes de *Sac. erythraea* avec le plasmide pXhoI.

La préparation des protoplastes et le processus d'intégration ont été réalisés comme à l'exemple 3.

La sélection et l'analyse des mutants ayant le phénotype
10 *ery*⁻ a été réalisée selon les conditions décrites à l'exemple 19.

L'intégration dans le chromosome et la présence de la délétion attendue ont été confirmées par analyse de Southern selon les méthodes générales décrites à l'exemple 19.

15 Après la transformation des protoplastes avec le plasmide pXhoI et la sélection des colonies résistantes au thiostrepton, l'intégration dans le chromosome a été confirmée par analyse de Southern. En utilisant comme sonde le fragment *Pst*I de 3,3 kb du plasmide pNCO62, l'ADN
20 chromosomique d'un intégrant préalablement digéré avec les enzymes de restriction *Pst*I et *Bgl*III a montré les bandes 3 kb et 6 kb attendues.

Après cultures répétées des intégrants, les colonies individualisées obtenues ont été analysées pour la sensi-
25 bilité au thiostrepton et la production d'érythromycine, puis l'intégration de la délétion attendue (délétion de 60 pb du nucléotide 47338 au nucléotide 47397 de la séquence de la figure 3) dans le chromosome d'un clone mutant ayant le phénotype *ery*⁻ (XhoI) a été confirmée par analyse de
30 Southern. L'ADN chromosomique, isolé respectivement à partir de la souche sauvage et du mutant XhoI, a été digéré avec l'enzyme de restriction *Pst*I. L'hybridation avec la sonde *Pst*I-NcoI de 0,8 kb (nucléotides 46368 à 47142 de la séquence de la figure 3) a donné le profil attendu avec une bande *Pst*I
35 de 1 kb correspondant aux nucléotides 46368 à 47397 de la séquence de la figure 3 à partir de la souche sauvage et avec une bande de 4 kb à partir du mutant indiquant que le site *Pst*I à la position 47397 ci-dessus a été perdu.

La perte du site *Pst*I à la position 47397 a aussi été confirmée par PCR. L'ADN chromosomique a été soumis à une amplification par PCR en utilisant les amorces correspondant respectivement à la séquence du nucléotide 47300 au nucléotide 57320 et à la séquence du nucléotide 47661 au nucléotide 47636 de la séquence de la figure 3. Un fragment attendu de 306 pb a été ainsi amplifié à partir de la souche sauvage générant après digestion avec l'enzyme de restriction *Pst*I deux bandes d'environ 100 et 300 pb. A partir du mutant Xho91, un fragment de 300 pb a été amplifié, résultant de la délétion de 60 pb. Ce fragment a été ensuite isolé et a été trouvé résistant à la digestion par l'enzyme *Pst*I.

La souche recombinante ainsi obtenue et désignée Xho91, a ensuite été cultivée pour identifier les métabolites produits.

EXEMPLE 23 : fermentation de la souche Xho91 et identification des métabolites secondaires produits.

La culture de la souche Xho91 et l'analyse du surnageant de culture par spectrométrie de masse ont été réalisées selon les conditions décrites à l'exemple 20.

La production d'érythromycine A (m/z 734 et m/z 716) n'a pas été observée mais la présence d'une quantité majoritaire d'érythronolide B (MK^+ : m/z 441 ; MNa^+ : m/z 425 ; $M-H_2O$ H^+ : m/z 385) ainsi que la présence de désosaminyl érythronolide B (m/z 560) mises en évidence caractérisent la souche *Pst*10 comme un mutant *eryB*.

Les résultats de spectrométrie de masse ont été confirmés par spectrométrie de masse en haute résolution sur un spectromètre Brucker FT-ICR (Brucker, FRG).

La séquence *eryBVI* présente une forte homologie avec *DnmT* impliquée dans la biosynthèse de la daunorubicine chez *S. peucetius* (numéro d'accèsion EMBL U77891) (43,9 % d'identité au niveau protéique).

Ces observations indiquent que le gène *eryBVI* code pour la dTDP-4-céto-L-6-désoxyhexose 2,3-déshydratase impliquée dans la biosynthèse du dTDP-mycarose, comme suggéré par Scotti et Hutchinson, 1996.

EXEMPLE 24 : construction du plasmide pCIVA.

Un plasmide d'intégration, dénommé pCIVΔ et portant une délétion dans le gène *eryCIV* codant pour l'ORF17, a été construit selon le schéma de la figure 13A de la façon suivante :

5 Le plasmide pNC062 obtenu à l'exemple 11 a été digéré à l'aide des enzymes de restriction *BalI* et *BclI* de façon à éliminer un fragment ayant 949 pb à l'intérieur de l'ORF17 du nucléotide 48650 au nucléotide 49598 de la séquence de la figure 3. Après remplissage des extrémités à l'aide du
10 fragment de Klenow de l'ADN polymérase I, le plasmide a été religaturé et transformé dans *E. coli* XL1-blue. A partir du plasmide pBCB17 ainsi généré, le fragment de 2,68 kb portant la délétion a été isolé par digestion à l'aide des enzymes *XbaI* et *SphI*, puis sous-cloné dans les sites correspondant du
15 plasmide pUWL218. La présence de la délétion de 949 pb du nucléotide 48650 au nucléotide 49598 de la séquence de la figure 3 a été confirmée par séquençage. Le plasmide d'intégration pCIVΔ ainsi obtenu (figure 13B) a ensuite été transféré dans la souche *E. coli* DH5αMRC, puis utilisé pour
20 transformer *Sac. erythraea*.

EXEMPLE 25 : construction d'une souche *Sac. erythraea eryCIVΔ* (CIV89).

Une souche, dans laquelle le gène *eryCIV* porte une délétion interne telle que celle introduite dans le plasmide
25 pCIVΔ obtenu à l'exemple 24, a été préparée par transformation des protoplastes de *Sac. erythraea* avec le plasmide pCIVΔ.

La préparation des protoplastes, le processus d'intégration et la sélection des mutants ayant le phénotype
30 *ery⁻* ont été réalisés comme à l'exemple 3.

De plus, la présence de la délétion attendue dans le chromosome (délétion de 949 bp du nucléotide 48650 au nucléotide 49598 de la séquence de la figure 3 a été confirmée par analyse génomique par Southern blot ainsi que
35 par PCR selon les conditions décrites à l'exemple 3.

Par hybridation Southern sur l'ADN génomique digéré par l'enzyme de restriction *NcoI*, en utilisant comme sonde l'oligonucléotide ayant la séquence suivante :

C4-R AGCGGCTTGATCGTGTGGACCAGTAC (SEQ ID N° 53)
correspondant au brin complémentaire de la région d'ADN
située de la position 49996 à la position 50022 de la
séquence de la figure 3, une bande de 6,2 kb à partir de la
souche sauvage et une bande de 5,2 kb à partir du mutant
CIV89 ont été détectées. Les résultats montrés à la figure 15
indiquent la présence dans le mutant d'une délétion d'environ
1 kb dans cette région du chromosome.

L'amplification par PCR a été effectuée en utilisant
l'oligonucléotide ayant la séquence suivante :
C4-S GGCCTATGTGGACTACGTGTTGAACGT (SEQ ID N° 54)
correspondant à la région d'ADN située de la position 48169 à
la position 48195 de la séquence de la figure 3 et
l'oligonucléotide C4-R ayant la séquence indiquée ci-dessus,
permettant d'encadrer par amplification PCR la région portant
la délétion interne à l'ORF17. L'analyse par amplification
par PCR a permis de détecter une bande de 1,8 kb dans la
souche sauvage et une bande de 900 pb dans le mutant CIV89 de
façon identique au signal obtenu avec le plasmide pCIVΔ. Les
résultats montrés à la figure 16 confirment que la délétion
d'environ 900 pb détectée par l'analyse Southern est
identique à celle portée par le plasmide pCIVΔ (949 pb).

La souche recombinante ainsi obtenue, désignée CIV89, a
été ensuite cultivée pour identifier les métabolites produits
par la souche.

**EXEMPLE 26 : fermentation de la souche CIV89 et identifica-
tion des métabolites secondaires produits.**

La culture de la souche CIV89 et les analyses par ccm
ont été réalisées selon les conditions indiquées à
l'exemple 4.

Les résultats de ccm (figure 17) montrent que la souche
CIV89 accumule préférentiellement du 3- α -mycarosyl érythrono-
lide B ainsi que de l'érythronolide B comme attendu d'un
mutant eryC.

Des métabolites mineurs de faibles mobilités ont été
également détectés, puis extraits et analysés par RP-HPLC
couplée à la spectrométrie de masse selon les conditions
utilisées à l'exemple 4.

Un métabolite mineur donne un pic parent à m/z 720 et des produits de déshydratation et de fragmentation à m/z 702, m/z 576 et m/z 174. Le pic 174 peut correspondre à la 4-hydroxydésosamine et le pic 576 au 4'-hydroxydésosaminyl
5 érythronolide B.

Ces résultats suggèrent que la différence de m/z de 16 comparée à l'érythromycine D (pic parent m/z 704) est portée par le sucre aminé. La structure proposée pour ce métabolite est la 4'-hydroxy érythromycine D.

10 Ces observations indiquent que l'enzyme est impliquée dans le retrait du groupement hydroxyle dans la voie de biosynthèse de l'érythromycine et que le gène *eryCIV* code pour la dTDP-6-désoxyhexose 3,4-déshydratase.

La souche CIV89 a été déposée à la Collection Nationale
15 de Cultures de Microorganismes (CNCM) INSTITUT PASTEUR, 25, Rue du Docteur Roux 75724 PARIS CEDEX 15 FRANCE, le 16 juillet 1997 sous le numéro I-1905.

EXEMPLE 27 : construction du plasmide pCVA .

Un plasmide d'intégration, dénommé pCVA et portant une
20 délétion dans le gène *eryCV* codant pour l'ORF18, a été construit selon le schéma de la figure 14A de la façon suivante :

Le fragment *BalI-BamHI* (3,48 kb), obtenu à partir du plasmide pNC062 préparé à l'exemple 11 par digestion avec les
25 enzymes de restriction *BalI* et *BamHI*, a été sous-cloné dans les sites *SmaI-BamHI* du vecteur pUC19. Du plasmide résultant pBAB18 (figure 5B), le fragment interne *ScaI* (1kb) a été ensuite délété par digestion avec l'enzyme *ScaI* pour générer une délétion de 1044 pb du nucléotide 49998 au nucléotide
30 51041 de la séquence de la figure 3 dans l'ORF18. Du plasmide pBABACV ainsi obtenu, le fragment portant la délétion a ensuite été réisolé à partir du polylinker de pUC19 par digestion avec les enzymes de restriction *HindIII* et *EcoRI*, puis sous-cloné dans le plasmide pUWL218. Le plasmide
35 d'intégration pCVA ainsi obtenu (figure 14B) a été transféré dans la souche *E. coli* DH5αMRC, puis utilisé pour transformer *Sac. erythraea*.

EXEMPLE 28 : construction d'une souche *Sac. erythraea eryCVA*

(CV90).

Une souche dans laquelle le gène *eryCV* porte une délétion interne telle que celle introduite dans le plasmide pCVA obtenu à l'exemple 27 a été préparée par transformation
5 des protoplastes de *Sac. erythraea* avec le plasmide pCVA.

La préparation des protoplastes, le processus d'intégration et la sélection des mutants ayant le phénotype *ery*⁻ ont été réalisés comme à l'exemple 3.

De plus, la présence de la délétion attendue dans le
10 chromosome (délétion de 1044 pb du nucléotide 49998 au nucléotide 51041) a été confirmée par analyse génomique par Southern blot ainsi que par PCR selon les conditions décrites à l'exemple 3.

Par hybridation Southern sur l'ADN génomique digéré par
15 l'enzyme de restriction *Nco*I, en utilisant comme sonde l'oligonucléotide ayant la séquence suivante :
C5-R AACGCCTCGTCCTGCAGCGGAGACACGAACA (SEQ ID N° 55)
correspondant au brin complémentaire de la région d'ADN située de la position 51229 à la position 51259 de la
20 séquence de la figure 3, une bande de 6,2 kb à partir de la souche sauvage et une bande de 5,1 kb à partir du mutant CV90 ont été détectées. Les résultats montrés à la figure 15 indiquent chez le mutant la présence d'une délétion d'environ 1,1 kb dans cette région du chromosome.

25 L'amplification par PCR a été effectuée en utilisant l'oligonucléotide ayant la séquence suivante :
C5-S TTCGCTCCCGATGAACACAACCTCGTA (SEQ ID N° 56)
correspondant à la région d'ADN située de la position 49668 à la position 49694 de la séquence de la figure 3 et
30 l'oligonucléotide C5-R ayant la séquence indiquée ci-dessus, permettant d'encadrer par amplification PCR la région portant la délétion interne à l'ORF18. L'analyse par amplification par PCR a permis de détecter une bande d'environ 1,6 kb dans la souche sauvage et une bande d'environ 500 pb dans le
35 mutant CV90 de façon identique au signal obtenu avec le plasmide pCVA. Les résultats montrés à la figure 16 confirment que la délétion d'environ 1,1 kb détectée par l'analyse Southern est identique à celle portée par le plasmide PCVA

(1044 pb).

La souche recombinante ainsi obtenue, désignée CV90, a été ensuite cultivée pour identifier les métabolites produits par la souche.

5 EXEMPLE 29 : fermentation de la souche CV90 et identification des métabolites secondaires produits.

La culture de la souche CV90 et les analyses par ccm ont été réalisées selon les conditions indiquées à l'exemple 4.

Les résultats de ccm (figure 17) montre que la souche
10 CV90 accumule préférentiellement du 3- α -mycarosyl érythronolide B ainsi que de l'érythronolide B comme attendu d'un mutant eryC.

La séquence des résidus 38-50 (VTGAGDGDADVQA)
Val Thr Gly Ala Gly Asp Gly Asp Ala Asp Val Gln Ala
15 (SEQ ID N° 61)
de la protéine codée par eryCV (séquence de SEQ ID N° 11) est proche de la séquence consensus de liaison au NAD⁺ décrit par Wierenga et al., 1985 et par Scrutton et al., 1990.

Ces observations permettent de conclure que le gène
20 eryCV code pour une réductase qui interviendrait comme une dTDP-4,6-désoxyhexose 3,4-réductase dans la voie de biosynthèse de la d-TDP-désosamine.

EXEMPLE 30 : surexpression du produit du gène eryCIII dans E. coli.

25 L'expression hétérologue du produit du gène eryCIII de *Sac. erythraea* correspondant à l'ORF8 décrite à l'exemple 1 et codant pour l'activité désosaminyltransférase identifiée à l'exemple 7 a été réalisée en utilisant *E. coli* comme souche hôte. La protéine ainsi produite sous forme de corps
30 d'inclusion a été ensuite purifiée et son activité enzymatique déterminée in vitro.

1) Expression de la protéine EryCIII dans E. coli

L'expression a été réalisée en utilisant le vecteur pET11a (Stratagène) pour le clonage et l'expression de
35 protéines recombinantes dans *E. coli* sous le contrôle du promoteur de l'ARN polymérase du bactériophage T7.

Dans un premier temps, le gène eryCIII a été amplifié à partir du plasmide pK62 décrit à l'exemple 1 de la façon

suivante :

L'amplification par PCR a été effectuée en utilisant la polymérase Native Pfu (Stratagène) et comme amorces l'oligonucléotide A homologue au brin codant du gène *eryCIII* 5 ayant la séquence
A GAAGGAGATATACATATGCGCGTCGTCTTCTCCTC (SEQ ID N° 57)
permettant d'introduire un site *NdeI* en amont de l'ATG initiateur de *eryCIII* et l'oligonucléotide B homologue au brin complémentaire du gène *eryCIII* ayant la séquence
10 B CGGGATCCTCATCGTGGTTCTCTCCTTCCTGC (SEQ ID N° 58)
permettant d'introduire un site *BamHI* en aval du codon stop du gène *eryCIII*.

L'ADN amplifié a été ensuite digéré par les enzymes de restriction *NdeI* et *BamHI*, puis le fragment *NdeI*-*BamHI* de 15 1,2 kb obtenu contenant la totalité du gène *eryCIII* a été ligaturé dans le vecteur d'expression pET11a (Stratagène) qui contient le gène β -lactamase de résistance à l'ampicilline, l'origine de répllication *ColE1* et le promoteur du gène de l'ARN polymérase T7 situé en amont du site de clonage *NdeI*,
20 préalablement digéré avec les enzymes de restriction *NdeI* et *BamHI*. Après ligation et transformation dans *E. coli* XL1-blue, le plasmide pECIII ainsi obtenu a été confirmé par carte de restriction et séquençage.

La souche d'*E. coli* BL21(DE3) de la trousse pET 25 (Stratagène) qui contient dans son ADN chromosomique le gène *lacI^q* et le promoteur *lacUV5* en amont du gène de l'ARN polymérase T7, a ensuite été transformée par le plasmide pECIII.5.

La souche transformée obtenue, dénommée BL21/pECIII, a 30 été cultivée en erlen de 50 ml à 37°C en milieu LBensemencé à $DO_{600}=0,1$ à partir d'une préculture, puis induite par l'isopropyl- β -D-thiogalactopyranoside (IPTG) 1mM à $DO_{600}=1$. Après 3 h 30 d'induction, 1 ml de culture a été prélevé et centrifugé, puis le culot bactérien obtenu a été dissout dans
35 240 μ l d'eau et 120 μ l de tampon d'échantillon SDS 3X (Tris-HCl 1M pH = 6,8 : 1,9 ml ; glycérol 3 ml ; β -mercaptoéthanol 1,5 ml ; SDS 20 % , 3 ml ; bleu de bromophénol 1 % pH = 7 : 0,3 ml ; H₂O qsq 10 ml). A partir de 15 μ l de la solution

obtenue, les protéines totales extraites ont été analysées par SDS-PAGE sur un gel à 10 % de polyacrylamide avec une coloration au bleu de Comassie.

La surexpression d'une protéine ayant un poids moléculaire apparent d'environ 46 Kd correspondant au PM attendu pour la protéine EryCIII a été observée comparative-ment aux protéines totales d'une souche témoin transformée par le plasmide pET11a.

2) Purification de la protéine EryCIII

La souche transformée BL21/pECIII ci-dessus a été cultivée en fermenteur de 6 litres en milieu minimum contenant du glycérol comme source de carbone (Korz et al., 1995) à 25°C jusqu'à $D_{0600}=12$, puis induite par l'IPTG pendant 18 h jusqu'à $D_{0600}=54$. A partir du bouillon récolté, le culot bactérien contenant des corps d'inclusion a été isolé par centrifugation à 5000 g pendant 30 mn.

L'induction de la protéine EryCIII a été contrôlée par SDS-PAGE (gradient de polyacrylamide : 10 à 20 %) et avec une coloration au bleu de Comassie après lyse sur un aliquot dans le tampon SDS 1 %, à 100°C pendant 5 min, soit directement sur le bouillon récolté, soit sur le culot bactérien après une première lyse par sonication dans un tampon phosphate.

190 g de culot bactérien correspondant à 1 litre de bouillon récolté ont été remis en suspension dans 2,5 volumes de tampon $\text{KH}_2\text{PO}_4/\text{K}_2\text{HPO}_4$ 20 mM pH 7,2 contenant de l'EDTA 2,5 mM et du DTT 2,5 mM. Les cellules ont été ensuite lysées en utilisant un appareil Rannie (Mini-Lab, type 8-30H, APV Homogenisers As, Denmark) avec trois passages sous une pression de 1000 bars. Après centrifugation à 46.000 g pendant 3 heures, le culot obtenu a été mis en suspension dans 2,5 volumes d'urée 2M puis centrifugé dans les mêmes conditions.

Le culot ainsi lavé a été ensuite mis en suspension dans 2,5 volumes d'une solution d'urée 7M dans du tampon tris 50 mM pH 7,5 (tampon A) de façon à solubiliser la protéine EryCIII. Après centrifugation dans les mêmes conditions, le surnageant recueilli obtenu contient 2,1 g de protéines totales déterminées par la méthode de Bradford en utilisant

une trousse du commerce (Pierce).

L'extrait dans l'urée 7M a été ensuite chargé à la vitesse de 0,5 mètres/h et à 4-8°C sur une colonne de 180 ml (5 cm x 9 cm) de Q sepharose (Pharmacia) préalablement
5 équilibrée avec le tampon A ci-dessus et avec une détection à 280 nm. La protéine EryCIII a été ensuite éluée avec le tampon A contenant NaCl 0,3M. Les fractions réunies, contenant la protéine EryCIII, mise en évidence par SDS-PAGE (gradient de polyacrylamide : 10 à 20 %) révélée par
10 coloration au bleu de Comassie et 835 mg de protéines totales, ont été ensuite chargées sur une colonne de 5,5 litres (10 cm x 70 cm) de Superdex 200 Prep grade (Pharmacia) préalablement équilibrée avec le tampon A ci-dessus. Par élution de la colonne avec le tampon A et par détection à
15 280 nm, un pic de protéine a été obtenu dont les fractions contenant la protéine EryCIII mise en évidence par SDS-PAGE et 200 mg de protéines totales ont été réunies puis purifiées sur une colonne de 180 ml (5 cm x 9 cm) de Q Source (Pharmacia) préalablement équilibrée avec le tampon A. Par
20 élution avec un gradient linéaire de NaCl variant de 0 à 0,3 M dans le tampon A, 30 ml de solution contenant 100 mg de protéine EryCIII dénaturée homogène en pureté évaluée par SDS-PAGE, avec une révélation au nitrate d'argent, ont été obtenus.

25 La figure 18 montre l'évolution de la pureté de la protéine EryCIII suivie par SDS-PAGE (gradient de polyacrylamide : 10 à 15 %) pour un dépôt de 500 ng de protéines totales et une révélation au nitrate d'argent successivement après extraction à l'urée 7M (ligne 2),
30 chromatographie Q sepharose (ligne 3), chromatographie Superdex (ligne 4), chromatographie Q source (ligne 6) par rapport aux marqueurs de poids moléculaire (lignes 1 et 5).

La protéine EryCIII a été ensuite renaturée par dilution de l'éluat homogène avec une solution de tampon A contenant
35 du DTT 10 mM pour obtenir une concentration finale en protéine de 0,1 mg/ml. La solution diluée a été ensuite dialysée contre du tampon Tris 50 mM ; NaCl 0,15 M ; 0,3 % n-octyl- β -D-glucopyranosyl (NOG) ; DTT 10 mM, pH 8,3 puis

concentrée à 4 mg/ml par ultrafiltration sur une membrane PLGC04310 (Millipore) ayant un seuil de coupure de 10.000.

La protéine EryCIII purifiée a été ensuite conservée à l'état congelé à -20°C en aliquots de 500 µl.

5 3) **Caractérisation de la protéine EryCIII**

La caractérisation de la protéine EryCIII ainsi obtenue a été examinée pour les propriétés suivantes :

a) Homogénéité.

L'électrophorèse par SDS-PAGE (gradient de
10 polyacrylamide : 10 à 15 %) en utilisant l'appareil Phast System (Pharmacia) et une révélation au nitrate d'argent montre une pureté supérieure à 99 % pour un dépôt de 2000 ng.

b) Poids moléculaire par électrophorèse et spectrométrie de masse.

15 Par électrophorèse, un PM apparent de 46 kDa a été déterminé en accord avec le PM calculé de 45929.

L'analyse par RP-HPLC couplée à la spectrométrie de masse (HPLC : ESI-SM) donne une masse de 45934 uma.

c) Séquence en acide aminés N-terminale

20 La séquence N-terminale a été déterminée par microséquençage sur un microséquenceur de protéine Model A492 couplé à un analyseur HPLC de PTH-aminoacides (Applied Biosystems).

Aucune séquence secondaire n'a été décelée pour les
25 10 premiers résidus qui est en accord avec la séquence en acides aminés décrite à la figure 2 (séquence de SEQ ID N° 5).

d) Activité biologique

L'activité désosaminyl transférase de la protéine
30 EryCIII a été déterminée in vitro par la mise en évidence de la formation d'érythromycine D à partir de dTDP-D-désosamine, dont la préparation est décrite plus loin et de 3- α -mycarosyl érythronolide B (MEB) dont la préparation est décrite ci-dessus dans Matériels et Méthodes générales.

35 Le milieu de réaction contient 150 nmoles de dTDP-D-désosamine, 137,4 nmoles de MEB et 1 mg de protéine EryCIII en utilisant les conditions opératoires suivantes :

Dans un tube en verre à vis, on introduit successivement

4,78 ml de tampon Tris 50 mM pH 7,3 (tampon B) ; 20 µl de dTDP-D-désosamine, sel de triéthylamine (150 nmoles) en solution dans le tampon B contenant EDTA 1 mM et PEFABLOC O, 4 mM (Merck) ; 100 µl de MEB (137,4 nmoles) en solution dans le tampon B et 1 mg de protéine EryCIII correspondant à 250 µl d'un aliquot de solution congelée obtenue ci-dessus.

Après homogénéisation au Vortex, le tube bouché est placé pendant 5 h dans un bain thermostaté à 30°C, puis on ajuste le pH à 9-10 avec NaOH 32 % puis extrait le mélange réactionnel 3 fois avec 5 ml d'acétate d'éthyle. L'extrait obtenu, amené à sec sous pression réduite, puis repris par 100 µl de chlorure de méthylène est ensuite analysé par ccm dans les conditions indiquées à l'exemple 4 en utilisant comme éluant le mélange chlorure de méthylène/méthanol (90 : 10, v/v).

Un essai témoin (t = 0) dont l'incubation est arrêtée immédiatement par l'ajout de NaOH, est effectué dans les mêmes conditions.

Les résultats obtenus par révélation chimique montrent l'apparition d'un produit moins mobile ayant un Rf voisin de l'érythromycine D attendue et pour lequel une faible activité antibiotique est détectée par autobiogramme direct des plaques sur *B. pumilus*. Aucune activité biologique n'est observée pour l'essai témoin (figure 19).

Ces résultats confirment que la protéine EryCIII produite dans *E. coli* et purifiée ci-dessus a l'activité désosaminyl transférase attendue et a été correctement renaturée.

EXEMPLE 31 : utilisation de la séquence du gène eryCIII comme sonde pour isoler les gènes oleg1 et oleg2 codant pour des glycosyltransférases chez *S. antibioticus*.

1) clonage des gènes oleg1 et oleg2

La séquence du gène eryCIII de *Sac. erythraea* correspondant à l'ORF8 décrite à l'exemple 1 codant pour l'activité désosaminyltransférase a été utilisée pour préparer une sonde d'hybridation et a permis d'isoler des gènes homologues dans la souche *S. antibioticus* ATCC 11891 productrice d'oléandomycine par hybridation Southern.

L'intégralité du gène *eryCIII* a été amplifiée par PCR à partir de 6 ng du plasmide pK62 obtenu à l'exemple 1 en suivant les conditions opératoires décrites à l'exemple 3 en utilisant la polymérase native pfu (Stratagene) et comme 5 amorces l'oligonucléotide ayant la séquence suivante :
eryCIII-1 CGGGTACCATGCGCGTCGTCTTCTCCTCCATG (SEQ ID N° 59) comportant un site de restriction *KpnI* dans sa région 5' et dont la partie 3' correspond au brin complémentaire de la région d'ADN située de la position 10196 à la position 10219 10 de la séquence de la figure 2 et l'oligonucléotide *eryCIII2* ayant la séquence suivante :

eryCIII-2 CGGGTACCTCATCGTGGTTCTCTCCTTCC (SEQ ID N° 60) comportant un site *KpnI* dans sa région 5' et dont la partie 3' correspond à la région d'ADN située de la position 8954 à 15 la position 8974 de la séquence de la figure 2.

La bande d'environ 1,2 kb obtenue par amplification a été ensuite digérée par l'enzyme de restriction *KpnI* et clonée dans le plasmide pUC19 préalablement digéré par l'enzyme *KpnI*. Le plasmide pCIIIPCR1 ainsi obtenu a été 20 ensuite utilisé pour réisoler le fragment *KpnI* de 1,2 kb correspondant à l'intégralité du gène *eryCIII* montré à la figure 2. Le fragment ainsi isolé a été ensuite marqué au ³²p par la technique "random priming" décrite par Sambrook et al., 1989 et utilisé comme sonde *eryCIII* pour analyser par 25 hybridation Southern des clones cosmides obtenus à partir d'une banque d'ADN génomique de *S. antibioticus* ATCC 11891 et préparés de la façon suivante (figure 20) :

Une série de six cosmides (cosAB35, cosAB76, cosAB87, cosAB67, cosAB63 et cosAB61) se chevauchant et couvrant 30 environ 100 kb de la région correspondant au cluster de gènes de la biosynthèse de l'oléandomycine a été isolée en suivant la méthode décrite par Swan et al., 1994 en utilisant comme sondes le fragment *SmaI* de 2 kb interne à la troisième sous-unité de la polykétide synthase de *Sac. erythraea* dans le 35 cluster de gènes de la biosynthèse de l'érythromycine (Cortes et al., 1990) suivie d'une marche sur le chromosome. Les sondes *strD*, *strE* et *strM* codant respectivement pour la dTDP-glucose synthase, la dTDP-glucose 4,6-déshydratase et la

dTDP-6-désoxyglucose 3,5-épimérase de *S. griseus* (Stockmann et Piepersberg, 1992) hybrident avec les cosmides cosAB61 et cosAB63 (fig. 20). De façon analogue, par hybridation Southern avec la sonde eryCIII préparée ci-dessus effectuée
5 selon les conditions standard décrites par Hopwood et al., 1985, le cosmide cosAB35 (Swan et al., 1994) donne des signaux positifs dans deux fragments de restriction *Bam*HI de 3,5kb et de 2,7 kb représentés à la figure 20. Le sous-clonage et le séquençage ultérieurs montrent que ces deux
10 fragments sont séparés par un fragment *Bam*HI de 0,6 kb non détecté par hybridation.

Un fragment *Sst*I de 10,8 kb d'ADN génomique de *S. antibioticus* ATCC 11891 représenté à la figure 21, correspondant à la partie droite du cluster de gènes de la
15 biosynthèse de l'oléandomycine comprise entre le site *Sst*I en position 11081 du gène OLE-ORF3 des PKS de la séquence EMBL n°L09654) et le site *Sst*I en position 5 de la séquence EMBL n°L36601 situé à 1,4 kb en amont du gène *oleB* et hybridant avec la sonde eryCIII préparée ci-dessus, a été isolé à
20 partir du cosmide cosAB35 et sous-cloné dans le vecteur plasmidique pSL1180 (Pharmacia Biotech). Le clone pC035-S ainsi obtenu a été utilisé pour générer des matrices simple brin par sous-clonage de différents fragments d'ADN dans les bactériophages M13mp18 et MP13mp19 (New England Biolabs),
25 puis la séquence nucléotidique de ces fragments a été déterminée selon la méthode de Sanger et al. (1977) en utilisant une polymérase T7 modifiée (Sequenase version 2.0; U.S. Biochemicals) en présence d' α [³⁵S]dCTP (Amersham) et de 7-déaza-dGTP, selon les recommandations du fournisseur afin
30 de limiter les problèmes de compression de bandes. Les amorces conventionnelles fournies avec la trousse Sequenase ainsi que les amorces synthétiques (17mer) internes ont été utilisées.

L'assemblage des données de séquence a été réalisé en
35 utilisant le programme Fragment Assembly (Genetic Computer Group, University of Wisconsin) et l'identification des phases ouvertes de lecture en utilisant le programme CODONPREFERENCE (Devereux et al., 1984).

Les séquences nucléotidiques obtenues ont permis d'établir la séquence nucléotidique de 6093 bp représentée à la figure 22 (séquence de SEQ ID N° 15), comprise entre les sites *SphI** et *KpnI* montrés à la figure 21, dans laquelle 5 cinq ORFs ont été identifiées respectivement du nucléotide 184 au nucléotide 1386 (ORF dénommée *oleP1*), du nucléotide 1437 au nucléotide 2714 (ORF dénommée *oleG1*), du nucléotide 2722 au nucléotide 3999 (ORF dénommée *oleG2*), du nucléotide 3992 au nucléotide 4729 (ORF dénommée *oleM*) et du nucléotide 10 4810 au nucléotide 5967 (ORF dénommée *oleY*). Les cinq ORFs sont transcrites dans la même direction.

Des échantillons de *E. coli* contenant le plasmide pCO35-S comprenant la région codante des ORFs *oleP1*, *oleG1*, *oleG2*, *oleM* et *oleY* ont été déposés à la Collection Nationale 15 de Cultures de Microorganismes (CNCM) INSTITUT PASTEUR, 25, Rue du Docteur Roux 75724 PARIS CEDEX 15 FRANCE, le 8 avril 1998 sous le numéro I-2003.

Le gène *oleG1* code pour un polypeptide ayant 426 acides aminés (séquence de SEQ ID N° 17). Cependant, la présence 20 d'un codon CGC codant pour une arginine très conservée dans cette classe de glycosyltransférase chez les Streptomycètes situé immédiatement en amont du codon GTG, indiquerait que le codon initiateur pourrait être le codon CTG en position 1431 de la séquence SEQ ID N° 17.

25 Le gène *oleG2* code pour un polypeptide ayant 426 acides aminés (séquence de SEQ ID N° 18).

La comparaison des séquences en acides aminés déduites des ORFs *oleG1* et *oleG2* ci-dessus avec les protéines de bases de données à l'aide du programme Blast (Altschul et al., 30 1990) a montré des similarités avec des glycosyl transférases de différentes sources, notamment environ 72 % de similarité et 53 % d'identité avec la déosaminylnyltransférase EryCIII décrite à l'exemple 30.

L'identification de la fonction du gène *oleG1* ou du gène 35 *oleG2* a été ensuite réalisée par interruption du gène cible dans la souche de *S. antibioticus* ATCC 11891 et par identification d'un précurseur non glycosylé de l'oléandomycine produit par la souche mutante en utilisant les méthodes

décrites dans les exemples 3 à 4.

2) génération d'une souche de *S. antibioticus* oleG1Δ (A35G1).

Une souche dans laquelle le gène *oleG1* est interrompu a été préparée par intégration d'un plasmide pCO3 dans les 5 régions homologues de l'ADN chromosomique de la souche de *S. antibioticus* ATCC 11891 productrice d'oléandomycine.

Dans un premier temps, le fragment *Bam*HI de 0,6 kb interne au gène *oleG1*, obtenu par digestion du plasmide pCO35-S préparé ci-dessus, avec l'enzyme de restriction *Bam*HI 10 (figure 21), a été sous-cloné dans le site *Bam*HI du plasmide pOJ260 NRRL B-14785.

Le plasmide pCO3 ainsi généré a été ensuite transféré dans la souche *E. coli* TG1 recO1504::Tn5 (Kolodner et al., 1985), puis utilisé pour transformer les protoplastes de 15 *S. antibioticus*. La sélection des transformants a été réalisée par résistance à l'apramycine (Apramycine pour injection, Rhône Mérieux).

La préparation des protoplastes a été réalisée à partir de la souche *S. antibioticus* ATCC 11891 en suivant les 20 conditions décrites par Hopwood et al., 1985.

La transformation a été réalisée en utilisant 50 µl d'aliquot de protoplastes, 5 µg d'ADN plasmidique pCO3 et en remplaçant le thiostrepton par de l'apramycine à la concentration finale de 25 µg/ml.

25 La sélection des intégrants effectuée par résistance à l'apramycine a permis d'isoler un clone dénommé A35G1.

L'altération attendue dans la région correspondante du chromosome de *S. antibioticus* a été confirmée par analyse génomique par Southern blot. L'ADN chromosomique isolé puis 30 digéré par l'enzyme de restriction *Pst*I à partir de la souche *S. antibioticus* sauvage ou du mutant A35G1 a été analysé par Southern en utilisant comme sonde d'hybridation le fragment *Bam*HI de 0,6 kb indiqué ci-dessus. Le remplacement du fragment *Pst*I de 4,7 kb ainsi détecté dans la souche sauvage 35 par deux fragments *Pst*I de 2,4 et 6,5 kb dans le mutant A35G1 confirme l'intégration du plasmide pCO3 dans le chromosome de la souche A35G1 au niveau de l'ORF *oleG1*.

La souche recombinante A35G1 ainsi obtenue a été ensuite

cultivée pour identifier les précurseurs produits par la souche.

3) fermentation de la souche A35G1 et identification des précurseurs de l'oléandomycine produits.

5 La souche A35G1 a été cultivée pendant 72 h en erlen de 50 ml dans le milieu EP2 à partir d'une préculture de 48 h en milieu EP1 dans les conditions décrites à l'exemple 4.

Les extraits de bouillon avec de l'acétate d'éthyle ont été ensuite analysés selon les méthodes utilisées dans les
10 exemples 3 et 4.

a) L'essai biologique par antibiogramme a été réalisé de la façon suivante :

Après croissance de la souche *B. pumilus* sur milieu TSB pendant une nuit à 37°C, la culture a été diluée à 1/100 dans
15 du milieu contenant 50 % (w/v) de glycérol, puis la suspension cellulaire obtenue a été conservée à -20°C avant utilisation.

L'essai biologique a ensuite été effectué en introduisant 150 µl de la suspension cellulaire décongelée
20 dans 100 ml de milieu TSB contenant 1 % d'agar et maintenu à 55°C. Le mélange a été ensuite versé dans des boîtes de pétri. Après refroidissement, des cylindres oxford contenant 50 à 200 µl d'extraits à l'acétate d'éthyle ont été placés sur les boîtes d'agar, maintenus 2 h à 4°C, puis incubés
25 pendant une nuit à 37°C.

Les extraits ne montrent pas d'effet inhibiteur sur la croissance de *B. pumilus* ATCC 14884.

b) L'analyse par ccm par révélation chimique a été effectuée selon les conditions décrites à l'exemple 4 en
30 utilisant comme standards l'érythromycine A, l'érythronolide B ainsi que le 6-désoxyérythronolide B.

L'analyse par ccm montre que la souche A35G1 ne produit pas d'oléandomycine mais accumule préférentiellement un produit pourpre ayant une mobilité plus grande que
35 l'érythronolide B et voisine du 6-désoxyérythronolide B et que l'on peut attendre de la partie aglycone 8,8a-désoxy-oléandolide.

c) L'analyse par chromatographie RP-HPLC couplée à la

spectrométrie de masse a été réalisée selon les conditions décrites à l'exemple 4. Deux métabolites majeurs, dénommés M9 et M10, ont été détectés (élution à 6,12 mn et 17,23 mn respectivement). Les deux produits donnent un pic parent 5 m/z 373 et des profils de fragmentation analogues qui peuvent être en accord avec la structure [8,8a-désoxyoléandolide] H^+ . Cependant seul le temps de rétention du métabolite M10 est en accord avec la structure proposée alors que le métabolite M9 pourrait correspondre à une structure isomère ou au noyau 10 lactone ouvert.

Des expériences de complémentation des souches mutantes de *Sac. erythraea* CIII68 décrite à l'exemple 6 et BV88 décrite à l'exemple 16 ont été également réalisées en utilisant des constructions plasmidiques permettant 15 d'exprimer respectivement chacun des gènes *oleG1* et *oleG2*.

Ces observations et l'absence de détection d'oléandrosyl 8,8a-désoxyoléandolide ou de désosaminy 8,8a-désoxyoléandolide indiquent que le gène *oleG1* code pour la désosaminosyl-transférase et le gène *oleG2* code pour l'oléandrosyltransfé- 20 rase respectivement impliquées dans la biosynthèse de l'oléandomycine.

PREPARATION DE L'EXEMPLE 30 : Thymidine 5'-(trihydrogen-diphosphate), P'-[3,4,6-tridéoxy-3-(diméthylamino)-D-xylohexopyranosyl]ester, N,N-diéthyléthanamine

25 **STADE A** : chlorhydrate 3,4,6-tridéoxy-3-(diméthylamino)-D-xylohexopyranose

On ajoute sous agitation et à température ambiante 146,6 g d'érythromycine A, à 1,5 litres d'acide chlorhydrique 6N. On porte la solution obtenue au reflux pendant 2 h. On 30 refroidit à la température ambiante, filtre et lave à l'eau le résidu obtenu. On extrait la phase aqueuse au chlorure de méthylène, puis à l'éther sulfurique. On ajoute à la phase aqueuse 10 g de noir L₂S et chasse les traces d'éther sous pression réduite. On filtre, et concentre. On effectue un 35 second envoi dans les mêmes conditions. On rassemble les deux essais, dissout dans l'éthanol (150 cm³), ajoute 150 cm³ d'éther éthylique. On essore, lave et sèche les cristaux obtenus. On obtient 42 g de produit recherché. F = 158-160°C.

STADE B : 3,4,6-tridéoxy-3-(diméthylamino)-D-xylo-hexopyranose,1,2-diacétate

On ajoute sous agitation à 20°C, 60 cm³ de triéthylamine dans un mélange renfermant 15,27 g de produit du stade A et 5 150 cm³ de chlorure de méthylène. On ajoute à 20°C une solution renfermant 20 cm³ d'anhydride acétique et 80 cm³ de chlorure de méthylène. On agite à la température ambiante pendant 20 heures. On filtre, lave et concentre. On empâte dans l'éther sulfurique. On concentre le filtrat sous 10 pression réduite. On chromatographie le résidu obtenu en éluant avec le mélange acétate d'éthyle-triéthylamine (95-5). On obtient 18,6 g de produit recherché que l'on utilise tel quel dans le stade suivant.

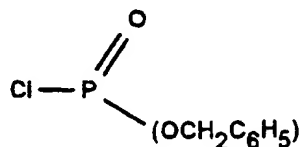
STADE C : 3,4,6-tridéoxy-3-(diméthylamino)-D-xylo-hexopyranose,2-acétate

On porte à 50°C un mélange renfermant 18,6 g du produit du stade B et 50 cm³ de DMF et ajoute 6,62 g d'acétate d'hydrazine NH₂NH₂, ACOH. On agite le mélange réactionnel et le verse sur une solution saturée de carbonate acide de 20 sodium. On extrait la phase aqueuse à l'acétate d'éthyle. On rassemble les phases organiques, sèche, filtre et concentre. On distille sous pression réduite pour éliminer le DMF par entraînement azéotrophique avec le toluène. On obtient 11,28 g de produit que l'on chromatographie sur silice en 25 éluant avec le mélange acétate d'éthyle-triéthylamine (90-10). On obtient 6,5 g de produit recherché que l'on utilise tel quel dans le stade suivant.

STADE D : 3,4,6-tridéoxy-3-(diméthylamino)-D-xylo-hexopyranose,2-acétate bis(phénylméthyl)phosphate

30 On ajoute à -70°C, 5,7 cm³ d'une solution de n-butyl-lithium dans l'hexane dans une solution renfermant 1,738 g du produit du stade précédent et 40 cm³ de THF. On ajoute à -70~-75°C 10 g de dibenzylphosphochlorure préparé extemporanément (J. Chem. Soc. 1958, p. 1957),

35



en solution dans 20 cm³ de THF. On maintient l'agitation pendant 1 h 30 entre -70 et -74°C. On verse sur une solution saturée de carbonate acide de sodium, extrait à l'acétate d'éthyle. On sèche la phase organique sur sulfate de sodium, 5 filtre et concentre. On chromatographie le produit obtenu sur silice en éluant avec le mélange acétone-chlorure de méthylène (5-5). On obtient 1,070 g du produit recherché.

STADE E : 3,4,6-tridéoxy-3-(diméthylamino)-D-xylo-hexopyranose,1-(dihydrogen phosphate),N,N-diéthyléthanamine

10 On place sous agitation et sous courant d'hydrogène pendant 30 minutes à la température ambiante, un mélange renfermant 1,070 g du produit du stade précédent, 20 cm³ d'acétate d'éthyle, 10 cm³ de méthanol, 0,622 cm³ de triéthylamine et 200 mg de palladium sur charbon. On filtre, lave au 15 méthanol et à l'acétate d'éthyle et concentre le filtrat. On obtient 1 g d'une huile à laquelle on ajoute 10 cm³ de méthanol. On agite la solution obtenue pendant 20 heures. On chasse le méthanol sous pression réduite à 30°C. On obtient 680 g de produit recherché.

20 STADE F : 3,4,6-tridéoxy-3-(diméthylamino)-D-xylo-hexopyranose,1-(dihydrogen phosphate)

On dissout 420 mg du produit isolé sous forme de sels de triéthylamine, obtenu au stade précédent, dans 1,6 cm³ de méthanol. On ajoute 3,2 cm³ d'éther sulfurique. On ajoute 25 ensuite 6,4 cm³ d'éther sulfurique. On obtient 250 mg de produit recherché fondant à 242-244°C.

STADE G : Thymidine 5'-(trihydrogen diphosphate),P'-[3,4,6-tridéoxy-3-(diméthylamino)-D-xylo-hexopyranosyl]ester,N,N-diéthyléthanamine

30 On mélange 228 mg du produit de la préparation 1, 6 cm³ de pyridine et 544 mg de thymidine 5'-monophosphate morpholide-4-morpholine-NN'-dicyclohexylcarboxamide. On chasse la pyridine sous pression réduite au rotorvapor en maintenant la température à 30°C ou en dessous. On ajoute 6 cm³ de pyridine 35 que l'on chasse sous pression réduite. On répète l'opération 2 fois. On ajoute 6 cm³ de pyridine, 105 mg de 1H-tétrazole. On agite pendant 3 jours à la température ambiante. On chasse la pyridine sous pression réduite. On reprend dans l'eau,

filtre, concentre et obtient un produit que l'on purifie par chromatographie. On obtient ainsi le produit recherché.

rf = 0,12 éluant CH₂Cl₂, MeOH, H₂O (60-35-6).

5 Références bibliographiques :

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990)
J Mol Biol 215 : 403-410.

10 Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA Struhl K (1995) Current protocols in molecular biology. Massachusetts General Hospital and Harvard medical School. John Wiley and Sons, Inc.

15 Bankier AT, Weston KM, Barrell BG (1987) Methods in Enzymology 155 : 51-93.

Bevitt DJ, Cortes J, Haydock SF and Leadlay PF, (1992), Eur. J. Biochem 204 : 39-49.

20

Caffrey P, Bevitt DJ, Staunton J, Leadlay PF (1992) FEBS 304 : 225-228.

Cortés J, Haydock SF, Roberts GA, Bevitt DJ, Leadlay PF
25 (1990) Nature 348 : 176-178.

Devereux J, Haeberli P, Smithies O (1984) Nucl Acids Res 12 : 387-395.

30 Dhillon N, Hale RS, Cortes J, Leadlay PF (1989) Mol Microbiol 3 : 1405-1414.

Dickens ML, Ye J, Strohl WR (1996) J Bacteriol 178 : 3384-3388.

35

Donadio S, Stassi J, McAlpine JB, Staver MJ, Sheldon PJ, Jackson M, Swanson SJ, Wendt-Pienkowski E, Wang YG, Jarvis B, Hutchinson CR, Katz L (1993) In : Baltz RH, Hegeman GD,

- Skatrud PL (eds) Industrial Microorganisms : Basic and Applied Molecular Genetics : American Society for Microbiology, Washington, DC, 257-265.
- 5 Gandecha AR, Large SL, Cundliffe E (1997) Gene 184 : 197-203.
- Haydock SF, Dowson JA, Dhillon N, Roberts GA, Cortes J, Leadlay PF (1991) Mol Gen Genet 230 : 120-128.
- 10 Hessler PE, Larsen PE, Constantinou AI, Schram KH, Weber JM, Appl Microbiol. Biotechnol (1997), 47 : 398-404.
- Hopwood DA, Kieser T, Wright HM and Bibb MJ, Journal of General Microbiology (1983), 129, 2257-2269.
- 15 Hopwood DA, Bibb MJ, Chater KF, Kieser T, Bruton C, Kieser HM, Lydiate DJ, Smith CP, Ward JM, Schremp H (1985) Genetic manipulation of Streptomyces. A laboratory Manual. John Innes Foundation, Norwich.
- 20 Kaneda T, Butte JC, Taubman B, Corcoran JW (1962) J Biol Chem 237 : 322-327.
- Katz E, Thompson CJ, Hopwood DA (1983) J Gen Microbiol 129 : 2703-2714.
- 25
- Korz DJ, Rinas U, Hellmuth K, Sanders EA, Deckwer W-D (1995) Journal of Biotechnology 39 : 59-65.
- Katz L, Donadio S (1995) Macrolides. In : Vining LC, Stuttard C (eds). Genetics and Biochemistry of Antibiotic Production. Butterworth-Heinemann, Newton, MA, 385-420.
- 30
- Liu H-W, Thorson JS (1994) Annu Rev Microbiol 48 : 223-56.
- 35 Otten SL, Liu X, Ferguson J, Hutchinson CR (1995) J Bacteriol 177 : 6688-6692.
- Sakakibara H and Omura S (1984). In : Omura S. (ed) Macrolide

Antibiotics : Chemistry, Biology, and Practice. Academic Press, Inc. London, 85-125.

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: 5 a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, N.Y.

Sanger F, Nicklen S, Coulson AR (1977) Proc Natl Acad Sci USA 74 : 5463-5467.

10

Scotti C, Hutchinson CR (1996) J Bacteriol 178 : 7316-7321.

Scrutton NS, Berry A, Perham RN (1990) Nature 343 : 38-43.

15 Staden R (1984), Nucleic Acids Res 12 : 521-528.

Stassi D, Donadio S, Staver MJ, Katz L (1993) J Bacteriol 175 : 182-189.

20 Stemmer WPC (1994) Proc. Natl. Acad. Sci., Vol. 91, pp. 10747-10751.

Stockmann M, and Piepersberg W (1992) FEMS Microbiol Lett 90 : 185-190.

25

Swan D.G., Rodriguez A.M., Vilches C., Méndez C., Salas J.A., (1994) Mol Gen Genet 242 : 358-362.

Ward JM, Janssen GR, Kieser T, Bibb MJ, Buttner MJ, Bibb MJ 30 (1986) Mol. Gen. Genet. 203 : 468-475

Weber JM, Wierman CK, Hutchinson CR (1985) J Bacteriol 164 : 425-433.

35 Weber JM, Losick R (1988) Gene 68 : 173-180.

Weber JM, Schoner B, Losick R (1989) Gene 75 : 235-241.

Weber JM, Leung JO, Maine GT, Potenz RHB, Paulus TJ, DeWitt JP (1990) J Bacteriol 172 : 2372-2383.

Weber JM, Leung JO, Swanson SJ, Idler KB, McAlpine JB (1991) 5 Science 252 : 114-117.

Wehmeier UF (1995) Gene 165 : 149-150.

Wierenga RK, Terpstra P, Hol WGJ (1986) J. Mol. Biol. 187 : 10 101-107.

Yamamoto H, Maurer KH, Hutchinson CR (1986) J Antibiot 34 : 1304-1313.

Texte libre de la liste de séquences

SEQ ID NO: 1:

/function= "implique dans la biosynthese du mycarose"

5 /gene= "eryBII"

/function= "implique dans la biosynthese de la desosamine"

/gene= "eryCII"

SEQ ID NO: 4:

10 /function= "implique dans la biosynthese de la desosamine"

/gene= "eryCIII"

/note= "SEQ ID NO 1 DE 1046 A 2308"

SEQ ID NO: 6:

15 /function= "implique dans la biosynthese du mycarose"

/gene= "eryBIV"

/transl_except= (pos: 242 .. 244, aa: Met)

/function= "implique dans la biosynthese du mycarose"

/gene= "eryBV"

20 /transl_except= (pos: 1210 .. 1212, aa: Met)

/function= "implique dans la biosynthese de la desosamine"

/gene= "eryCVI"

/function= "implique dans la biosynthese du mycarose"

/gene= "eryBVI"

25 /transl_except= (pos: 3308 .. 3310, aa: Met)

/function= "implique dans la biosynthese de la desosamine"

/gene= "eryCV"

/function= "implique dans la biosynthese du mycarose"

/gene= "eryBVII"

30 /transl_except= (pos: 7578 .. 7580, aa: Met)

SEQ ID NO: 13:

/function= "implique dans la biosynthese de la desosamine"

/gene= "eryCIV"

35 /note= "SEQ ID NO 6 DE 4837 A 6039"

SEQ ID NO: 15:

/gene= "oleP1"

```
/function= "glycosylation de 8,8a-desoxyoleandolide"  
/gene= "oleG1"  
/transl_except= (pos: 1437 .. 1439, aa: Met)  
/function= "glycosylation de 8,8a-desoxyoleandolide"  
5 /gene= "oleG2"  
/gene= "oleY"
```

```
SEQ ID NO: 20:  
/gene= "oleM"  
10 /note= "SEQ ID NO 15 DE 3992 A 4729"
```

```
SEQ ID NO: 22 à SEQ ID NO: 60  
/desc = "OLIGONUCLEOTIDE"
```

```
15 SEQ ID NO: 61:  
/note= "SEQ ID NO 11 DE 38 A 50"
```


REVENDEICATIONS

- 1) Séquence d'ADN simple ou double brin isolée, représentée à la figure 2 (séquence directe ou complémentaire de SEQ ID N° 1) correspondant à la région eryG-eryAIII du cluster de 5 gènes de la biosynthèse de l'érythromycine.
- 2) Séquence d'ADN selon la revendication 1 comprenant :
 - la séquence eryBII correspondant à l'ORF7 (séquence complémentaire de SEQ ID N° 1 du nucléotide 48 au nucléotide 1046) et codant pour une dTDP-4-céto-L-6-désoxyhexose
 - 10 2,3-réductase,
 - la séquence eryCIII correspondant à l'ORF8 (séquence complémentaire de SEQ ID N° 1 du nucléotide 1046 au nucléotide 2308) et codant pour une désosaminyltransférase et
 - la séquence eryCII correspondant à l'ORF9 (séquence complémentaire de SEQ ID N° 1 du nucléotide 2322 au nucléotide
 - 15 3404) et codant pour une dTDP-4-céto-D-6-désoxyhexose 3,4-isomérase.
- 3) Séquence d'ADN isolée représentée à la figure 2 choisie parmi la séquence eryBII correspondant à l'ORF7 (séquence
- 20 complémentaire de SEQ ID N° 1 du nucléotide 48 au nucléotide 1046), la séquence eryCIII correspondant à l'ORF8 (séquence complémentaire de SEQ ID N° 1 du nucléotide 1046 au nucléotide 2308) ou la séquence eryCII correspondant à l'ORF9 (séquence complémentaire de SEQ ID N° 1 du nucléotide 2322 au
- 25 nucléotide 3404) et les séquences qui hybrident et/ou présentent des homologues significatives avec cette séquence ou des fragments de celle-ci et ayant la même fonction.
- 4) Séquence d'ADN isolée eryCIII représentée à la figure 2 correspondant à l'ORF8 (séquence complémentaire de SEQ ID
- 30 N° 1 du nucléotide 1046 au nucléotide 2308 = séquence complémentaire de SEQ ID N° 4) et codant pour une désosaminyl transférase.
- 5) Polypeptide codé par l'une des séquences d'ADN selon l'une des revendications 1 à 4.
- 35 6) Polypeptide selon la revendication 5 correspondant à une ORF représentée à la figure 2, choisie parmi l'ORF7 (ayant la séquence de SEQ ID N° 2), l'ORF8 (ayant la séquence de SEQ ID N° 5) ou l'ORF9 (ayant la séquence de SEQ ID N° 3) et les

analogues de ce polypeptide.

7) Polypeptide selon la revendication 5 correspondant à l'ORF8 représentée à la figure 2 (ayant la séquence de SEQ ID N° 5) et ayant une activité désosaminyltransférase, dénommée 5 EryCIII.

8) Séquence d'ADN isolée représentée à la figure 3 (séquence de SEQ ID N° 6) correspondant à la région eryAI-eryK du cluster de gènes de la biosynthèse de l'érythromycine.

9) Séquence d'ADN selon la revendication 8 comprenant :

- 10 - la séquence eryBIV correspondant à l'ORF13 (séquence de SEQ ID N° 6 du nucléotide 242 au nucléotide 1207) et codant pour une dTDP-4-céto-L-6-désoxyhexose 4-réductase,
- la séquence eryBV correspondant à l'ORF14 (séquence de SEQ ID N° 6 du nucléotide 1210 au nucléotide 2454) et codant pour 15 une mycarosyltransférase,
- la séquence eryCVI correspondant à l'ORF15 (séquence de SEQ ID N° 6 du nucléotide 2510 au nucléotide 3220) et codant pour une dTDP-D-6-désoxyhexose 3-N-méthyltransférase,
- la séquence eryBVI correspondant à l'ORF16 (séquence de SEQ 20 ID N° 6 du nucléotide 3308 au nucléotide 4837) et codant pour une dTDP-4-céto-L-6-désoxyhexose 2,3-déshydratase,
- la séquence eryCIV correspondant à l'ORF17 (séquence de SEQ ID N° 6 du nucléotide 4837 au nucléotide 6039) et codant pour une dTDP-D-6-désoxyhexose 3,4-déshydratase,
- 25 - la séquence eryCV correspondant à l'ORF18 (séquence de SEQ ID N° 6 du nucléotide 6080 au nucléotide 7546) et codant pour une dTDP-D-4,6-didésoxyhexose 3,4-réductase et
- la séquence eryBVII correspondant à l'ORF19 (séquence de SEQ ID N° 6 du nucléotide 7578 au nucléotide 8156) et codant 30 pour une dTDP-4-céto-D-6-désoxyhexose 3,5 épimérase.

10) Séquence d'ADN isolée représentée à la figure 3 choisie parmi la séquence eryBIV correspondant à l'ORF13 (séquence de SEQ ID N° 6 du nucléotide 242 au nucléotide 1207), la séquence eryBV correspondant à l'ORF14 (séquence de SEQ ID 35 N° 6 du nucléotide 1210 au nucléotide 2454), la séquence eryCVI correspondant à l'ORF15 (séquence de SEQ ID N° 6 du nucléotide 2510 au nucléotide 3220), la séquence eryBVI correspondant à l'ORF16 (séquence de SEQ ID N° 6 du nucléo-

- tide 3308 au nucléotide 4837), la séquence eryCIV correspondant à l'ORF17 (séquence de SEQ ID N° 6 du nucléotide 4837 au nucléotide 6039), la séquence eryCV correspondant à l'ORF18 (séquence de SEQ ID N° 6 du nucléotide 6080 au nucléotide 7546) ou la séquence eryBVII correspondant à l'ORF19 (séquence de SEQ ID N° 6 du nucléotide 7578 au nucléotide 8156) et les séquences qui hybrident et/ou présentent des homologues significatives avec cette séquence ou des fragments de celle-ci et ayant la même fonction.
- 10 11) Séquence d'ADN isolée eryBV représentée à la figure 3 correspondant à l'ORF14 (séquence de SEQ ID N° 6 du nucléotide 1210 au nucléotide 2454) et codant pour une mycarosyltransférase.
- 12) Polypeptide codé par l'une des séquences d'ADN selon
15 l'une des revendications 8 à 11.
- 13) Polypeptide selon la revendication 12 correspondant à une ORF représentée à la figure 3, choisie parmi l'ORF13 (ayant la séquence de SEQ ID N° 7), l'ORF14 (ayant la séquence de SEQ ID N° 8), l'ORF15 (ayant la séquence de SEQ ID N° 9),
20 l'ORF16 (ayant la séquence de SEQ ID N° 10), l'ORF17 (ayant la séquence de SEQ ID N° 14), l'ORF18 (ayant la séquence de SEQ ID N° 11) ou l'ORF19 (ayant la séquence de SEQ ID N° 12) et les analogues de ce peptide.
- 14) Polypeptide selon la revendication 12 correspondant à
25 l'ORF14 représentée à la figure 3 (ayant la séquence de SEQ ID N° 8) et ayant une activité mycarosyltransférase, dénommé EryBV.
- 15) Utilisation d'au moins l'une des séquences d'ADN choisie parmi les séquences eryBII (séquence complémentaire de SEQ ID N° 1 du nucléotide 48 au nucléotide 1046), eryCIII (séquence complémentaire de SEQ ID N° 1 du nucléotide 1046 au nucléotide 2308) ou eryCII (séquence complémentaire de SEQ ID N° 1 du nucléotide 2322 au nucléotide 3404) représentées à la figure 2, eryBIV (séquence de SEQ ID N° 6 du nucléotide 242
35 au nucléotide 1207), eryBV (séquence de SEQ ID N° 6 du nucléotide 1210 au nucléotide 2454), eryCVI (séquence de SEQ ID N° 6 du nucléotide 2510 au nucléotide 3220), eryBVI (séquence de SEQ ID N° 6 du nucléotide 3308 au nucléotide

4837), eryCIV (séquence de SEQ ID N° 6 du nucléotide 4837 au nucléotide 6039), eryCV (séquence de SEQ ID N° 6 du nucléotide 6080 au nucléotide 7546) ou eryBVII (séquence de SEQ ID N° 6 du nucléotide 7578 au nucléotide 8156) représentée à la 5 figure 3, pour synthétiser des métabolites secondaires hybrides chez *Sac. erythraea*.

16) Utilisation d'au moins l'une des séquences d'ADN choisie parmi les séquences eryBII (séquence complémentaire de SEQ ID N° 1 du nucléotide 48 au nucléotide 1046), eryCIII (séquence 10 complémentaire de SEQ ID N° 1 du nucléotide 1046 au nucléotide 2308) ou eryCII (séquence complémentaire de SEQ ID N° 1 du nucléotide 2322 au nucléotide 3404) représentées à la figure 2, eryBIV (séquence de SEQ ID N° 6 du nucléotide 242 au nucléotide 1207), eryBV (séquence de SEQ ID N° 6 du 15 nucléotide 1210 au nucléotide 2454), eryCVI (séquence de SEQ ID N° 6 du nucléotide 2510 au nucléotide 3220), eryBVI (séquence de SEQ ID N° 6 du nucléotide 3308 au nucléotide 4837), eryCIV (séquence de SEQ ID N° 6 du nucléotide 4837 au nucléotide 6039), eryCV (séquence de SEQ ID N° 6 du nucléotide 6080 au nucléotide 7546) ou eryBVII (séquence de SEQ ID N° 6 du nucléotide 7578 au nucléotide 8156) représentée à la 20 figure 3 ou d'un fragment de cette séquence, comme sondes d'hybridation.

17) Utilisation de la séquence d'ADN eryCIII représentée à la 25 figure 2 (séquence complémentaire de SEQ ID N° 1 du nucléotide 1046 au nucléotide 2308 = séquence complémentaire de SEQ ID N° 4) comme sonde d'hybridation pour isoler des gènes homologues responsables de la glycosylation de la macrolactone chez une souche productrice de macrolide.

30 18) Utilisation selon la revendication 17 dans laquelle les gènes homologues sont les gènes de la biosynthèse de l'oléandomycine chez *S. antibioticus*.

19) Séquence d'ADN isolée représentée à la figure 22 (séquence de SEQ ID N° 15) correspondant à une région du 35 cluster de gènes de la biosynthèse de l'oléandomycine comprenant :

- la séquence correspondant à l'ORF oleP1 du nucléotide 184 au nucléotide 1386,

- la séquence correspondant à l'ORF oleG1 du nucléotide 1437 au nucléotide 2714 codant pour une activité glycosyltransférase,
- la séquence correspondant à l'ORF oleG2 du nucléotide 2722 au nucléotide 3999 codant pour une activité glycosyltransférase,
- la séquence correspondant à l'ORF oleM du nucléotide 3992 au nucléotide 4720 (= séquence de SEQ ID N° 20) et
- la séquence correspondant à l'ORF oleY du nucléotide 4810 au nucléotide 5967.

20) Séquence d'ADN isolée représentée à la figure 22 choisie parmi la séquence correspondant à l'ORF oleG1 (séquence de SEQ ID N° 15 du nucléotide 1437 au nucléotide 2714 codant pour une activité glycosyltransférase et la séquence correspondant à l'ORF oleG2 (séquence de SEQ ID N° 15 du nucléotide 2722 au nucléotide 3999) codant pour une activité glycosyltransférase.

21) Séquence d'ADN isolée selon la revendication 20 correspondant à l'ORF oleG1 (séquence de SEQ ID N° 15 du nucléotide 1437 au nucléotide 2714) codant pour une activité désosaminyltransférase.

22) Séquence d'ADN isolée selon la revendication 20 correspondant à l'ORF oleG2 (séquence de SEQ ID N° 15 du nucléotide 2722 au nucléotide 3999) codant pour une activité oléandrosyltransférase.

23) Polypeptide codé par la séquence d'ADN correspondant à l'ORF oleG1 et ayant une activité désosaminyltransférase (séquence de SEQ ID N° 17).

24) Polypeptide codé par la séquence d'ADN correspondant à l'ORF oleG2 et ayant une activité oléandrosyltransférase (séquence de SEQ ID N° 18).

25) Procédé de préparation de métabolites secondaires hybrides chez *Sac. erythraea* dans lequel :

- on isole une séquence ADN contenant au moins une séquence eryB ou une séquence eryC du cluster de gènes de la biosynthèse de l'érythromycine représentée à la figure 2 (séquence complémentaire de SEQ ID N° 1) ou à la figure 3 (séquence de SEQ ID N° 6),

- on crée une modification dans la dite séquence et on obtient une séquence altérée,
 - on intègre la séquence altérée dans le chromosome de la souche hôte et on obtient une souche modifiée,
- 5 - on cultive la souche modifiée dans des conditions permettant la formation du métabolite secondaire hybride et
- on isole le métabolite secondaire hybride.
- 26) Procédé selon la revendication 25 dans lequel la séquence ADN code pour l'une des enzymes choisie parmi une
- 10 - dTDP-4-céto-L-6-désoxyhexose 2,3-réductase,
- désosaminyltransférase,
 - dTDP-4-céto-D-6-désoxyhexose 3,4-isomérase,
 - dTDP-4-céto-L-6-désoxyhexose 4-réductase,
 - mycarosyltransférase,
- 15 - dTDP-D-6-désoxyhexose 3-N-méthyltransférase,
- dTDP-4-céto-L-6-désoxyhexose 2,3-déshydratase,
 - dTDP-D-6-désoxyhexose 3,4-déshydratase,
 - dTDP-D-4,6-didésoxyhexose 3,4-réductase ou
 - dTDP-4-céto-D-6-désoxyhexose 3,5 épimérase.
- 20 27) Procédé selon la revendication 25 dans lequel l'altération de la séquence résulte dans l'inactivation d'au moins l'une des enzymes choisie parmi une
- dTDP-4-céto-L-6-désoxyhexose 2,3-réductase,
 - désosaminyltransférase,
- 25 - dTDP-4-céto-D-6-désoxyhexose 3,4-isomérase,
- dTDP-4-céto-L-6-désoxyhexose 4-réductase,
 - mycarosyltransférase,
 - dTDP-D-6-désoxyhexose 3-N-méthyltransférase,
 - dTDP-4-céto-L-6-désoxyhexose 2,3-déshydratase,
- 30 - dTDP-D-6-désoxyhexose 3,4-déshydratase,
- dTDP-D-4,6-didésoxyhexose 3,4-réductase ou
 - dTDP-4-céto-D-6-désoxyhexose 3,5 épimérase.
- 28) Procédé selon la revendication 27 dans lequel l'enzyme inactivée est une dTDP-4-céto-L-6-désoxyhexose 4-réductase.
- 35 29) Procédé selon la revendication 27 dans lequel l'enzyme inactivée est une dTDP-D-6-désoxyhexose 3,4-déshydratase.
- 30) Procédé selon la revendication 27 dans lequel l'enzyme inactivée est une mycarosyltransférase.

- 31) Procédé selon la revendication 27 dans lequel l'enzyme inactivée est une dTDP-4-céto-L-6-désoxyhexose 2,3-réductase.
- 32) Procédé selon la revendication 25 dans lequel le métabolite secondaire hybride isolé est un analogue de
5 l'érythromycine choisi parmi la 4"-céto-érythromycine, la 4'-hydroxy-érythromycine ou la 3"-C-désméthyl-2",3"-ène-érythromycine.
- 33) Procédé selon la revendication 25 dans lequel le métabolite secondaire hybride isolé est le désosaminyl
10 érythronolide B.
- 34) Souche de *Sac. erythraea* modifiée dans laquelle au moins l'une des enzymes choisie parmi une
- dTDP-4-céto-L-6-désoxyhexose 2,3-réductase,
 - désosaminyltransférase,
 - 15 - dTDP-4-céto-D-6-désoxyhexose 3,4-isomérase,
 - dTDP-4-céto-L-6-désoxyhexose 4-réductase,
 - mycarosyltransférase,
 - dTDP-D-6-désoxyhexose 3-N-méthyltransférase,
 - dTDP-4-céto-L-6-désoxyhexose 2,3-déshydratase,
 - 20 - dTDP-D-6-désoxyhexose 3,4-déshydratase,
 - dTDP-D-4,6-didésoxyhexose 3,4-réductase ou
 - dTDP-4-céto-D-6-désoxyhexose 3,5 épimérase
- est inactivée et produisant au moins un métabolite secondaire hybride.
- 25 35) Souche de *Sac. erythraea* modifiée (BII92) dans laquelle une dTDP-4-céto-L-6-désoxyhexose 2,3-réductase est inactivée et produisant la 3"-C-désméthyl-2",3"-ène-érythromycine C.
- 36) Souche de *Sac. erythraea* modifiée (BIV87) dans laquelle une dTDP-4-céto-L-6-désoxyhexose 4-réductase est inactivée et
30 produisant la 4"-céto-érythromycine.
- 37) Souche de *Sac. erythraea* modifiée (CIV89) dans laquelle une dTDP-D-6-désoxyhexose 3,4-déshydratase est inactivée et produisant la 4'-hydroxyérythromycine D.
- 38) Souche de *Sac. erythraea* modifiée (BV88) dans laquelle
35 une mycarosyltransférase est inactivée et produisant du désosaminyl érythronolide B.
- 39) Procédé de préparation de précurseurs de l'oléandomycine chez *S. antibioticus* dans lequel

- on crée une altération de la séquence du gène choisie parmi la séquence d'ADN correspondant à l'ORF oleG1 (séquence de SEQ ID N° 15 du nucléotide 1437 au nucléotide 2714) et la séquence d'ADN correspondant à l'ORF oleG2 (séquence 5 de SEQ ID N° 15 du nucléotide 2722 au nucléotide 3999) dans le chromosome d'une souche hôte et obtient une souche modifiée,
 - on cultive la souche modifiée dans des conditions permettant l'accumulation des précurseurs de l'oléandomycine 10 et
 - on isole ces précurseurs.
- 40) Procédé selon la revendication 39 dans lequel l'altération est créée dans la séquence d'ADN correspondant à l'ORF oleG1 (séquence de SEQ ID N° 15 du nucléotide 1437 au 15 nucléotide 2714) et dont il résulte au moins l'élimination de l'activité désaminyltransférase et l'accumulation du précurseur de l'oléandomycine 8,8a-désoxyoléandolide.
- 41) Thymidine 5'-(trihydrogène diphosphate), P'-[3,4,6-tridésoxy-3-(diméthylamino)-D-.xylo.-hexopyranosyl] ester 20 (dTDP-D-désosamine) et les sels d'addition avec les bases.

1/60

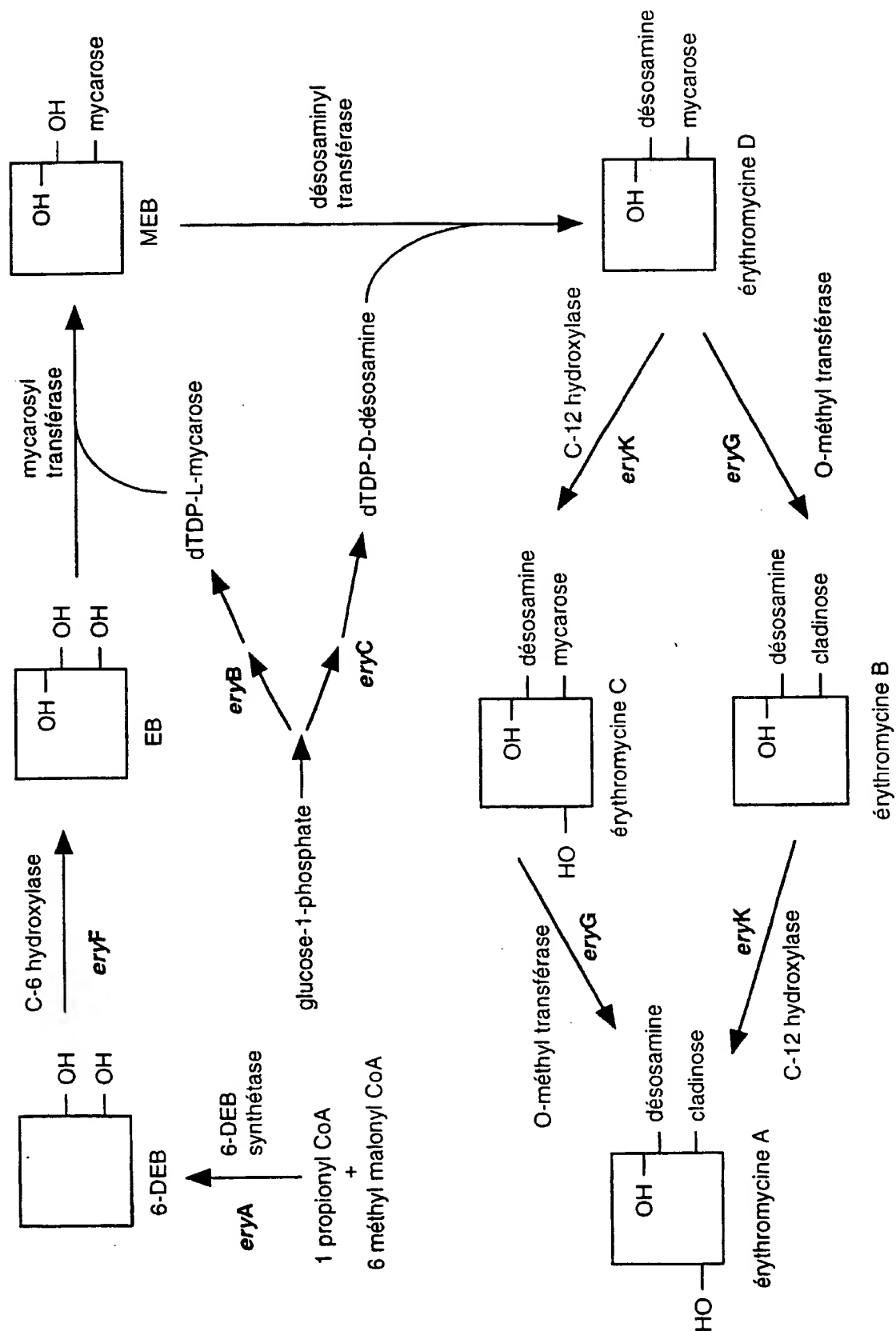


FIGURE 1

CTGCTTACGCTCACCAGCCGATCCTTTCTCGGTTCTTGTGCTCACTGCAACCAGG 7969
 +-----+-----+-----+-----+-----+-----+-----+-----+
 GACGAAGTGGAGTGGTCGGCATAGGAAAGAGCCAAAGGAGAACACGAGTACGTTGGTCC
 Q K V S V * Q L W
 <---ORF6

 CTTCGGCGCGCGCGCGCGGAGGCCACCGGGGGGAAGATCTCGTCCAGTTCGGACAGCG 8029
 +-----+-----+-----+-----+-----+-----+-----+-----+
 GAAGCGCGCGCGCGCGCGCTCCGGTGGCGCCCCCTTCTAGACGAGGTCAAGCCTGTGCGC
 A E P A A G G S A V A P F I E D L E S L

 CCTGCTCGTCCAGGGTCATCGCGGACGCCCTTACGCGGGAGTCGAGCTGCTCGGGGTTTC 8089
 +-----+-----+-----+-----+-----+-----+-----+-----+
 GGACGAGCAGGTCCAGTAGCGCCTGCGGAAGTCGCGCCTCAGCTCGACGAGCCCCCAAG
 A Q E D L T M A S A K L A S D L Q E P T

 GCGGGCCGATGACGGCGCGCGGCGATGCCGGGCGCGGACAGCACCCATGCGAGCCCCACCT 8149
 +-----+-----+-----+-----+-----+-----+-----+-----+
 CGCCCGGCTACTGCCGCGCGCGCTACGGCCCGCCCTGTCTGTGGGTACGCTCGGGGTGGA
 R P G I V A G A I G P R S L V W A L G V

 CGGCGGGGTCCTTCGCCGAGGTTGGGGCAGAACTTCTCGTAGGCCTCGATCGCCGGCGCA 8209
 +-----+-----+-----+-----+-----+-----+-----+-----+
 GCCGGCCAGAGCGGCTCCAACGCCGCTTGAAGAGCATCCGGAGCTAGCGGCCCGCGGT
 E A P D E G L N R C F K E Y A E I A P R

 GGGACGGCAACAGCACCTGCGCACGGCCCTGCCCGGACTTCACCGCGGTGCCCGCGGCCA 8269
 +-----+-----+-----+-----+-----+-----+-----+-----+
 CCCTGCCGTTGTGTTGACCGCTGCCGGGACCGCGCTGAAGTGGCGCCACGGGCGCGCGT
 L S P L L V Q A R G Q A S K V A T G A A

 GCTTCTCCAGCGCTCCGCTGAGCAGGCCCGCGTGTGAGCGGCGACCGGGAAGACGCCGA 8329
 +-----+-----+-----+-----+-----+-----+-----+-----+
 CGAAGAGGTCGCGAGCGGACTCGTCCGGGCGGACGTCGCCGCTGGTCCGCTTCTGCGGCT
 L K E L A G S L L G G H L P S W A F V G

FIGURE 2

3/60

```

8330  +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      GCCCGTAGGCCCTGCGCGGGCGGACACCTCCAGCTCGGCGTGCCGGACCGCCAGGTTGT      8389
      CGGCGATCCGGACGCGCGCCCGTCTGAGGTCGAGCCGACGGCCTGGCGGTCCAACA
      L G Y A Q A A P L V E L E A H R V A L N
8390  +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      ACAGGCACGTGGTGGAGACCATGCCACAGGAGTGCGCGGGCGGCGTTCTCTCTGCGCGG
      TGTCCGTGACCAACCTCTGGTACGGGTCCCTCACCGCCCGCCCGCAAGAGGACGCGCC      8449
      Y L C Q H S V M G L S H R R A A N E Q A
8450  +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      CGCGGATGTGCCAGCCCGCGAAGTTCGACGAGCCGACGTAGGAGACCTTGCCCGCTGGCGA
      GCCGCTACACGGTCGGCGCTTCAAGCTGCTCGGCTGCATCCTCTGGAACGGCGACCGCT      8509
      A A I H W G A F N S S G V Y S V K G S A
8510  +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      CGAGGCTGTCCATGGCCTGCCACACCTCGTCCACGGCGGACCGGTCGATGTGTGCA
      GCTCCGACAGGTACCGGACGGTGTGGAGCAGGTGCGCGCCTGGCCAGCTACACACGT      8569
      V L S D M A Q W V E D W P A S R D I H H
8570  +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      TCTGTAGACGTGATGTGGTCGACGCCCGCCAGCCTGCGCAGCGATCCCTCGCAGGAGGCGA
      AGACCATCTGCAGCTACACAGCTGCGGGTCGGACGCTCGCTAGGGAGCGTCTCCGCT      8629
      M Q Y V D I H D V G L R R L S G E C S A
8630  +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      TGATGTCCGCGCGACAGCCCGCTGTCTGTGACGCGCTCGCTCATCTCGCCGCCGACCT
      ACTACACGGCGGGCTGTGCGGCGACAGCAACTGCGCGAGCGAGTAGAGCGGCGGTGGA      8689
      I I H R A S L G S D N V R E S M E G G V
8690  +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      TGGTCGCCAGCACGGTGTCTCGCGCGCTCGCGCCCGCTGGGCCAGCCACCTGCCCCACCA
      ACCAGCGGTGTCACAGGAGCGGCGGACCGCGGGACCGGTCGGTGGACGGGTGCT      8749
      K T A L V T D E R R G G G Q A L W R G V

```

FIGURE 2

```

8750 GCTCTCGGTGTGGCCCTTGTAGAGCCGCCAGCCGTACATGTCGGCGGTGTCGAGGCAGT 8809
      +-----+-----+-----+-----+-----+-----+-----+-----+
      CGAGAGCCACACCGGGAACATCTCGGCGGTGCGCATGTACAGCCGCCACAGCTCCGTCA
      L E E T H G K Y L R W G Y M D A T D L C
8810 TGATGCCCGCGGTCCCGGGGTGGTCCATCAGGCGCAGCGGTGTCGTCCTCGACGCGTC 8869
      +-----+-----+-----+-----+-----+-----+-----+-----+
      ACTACGGCGCCAGGGCCCGCACAGGTAGTCCGCGTTCGCGCAGCAGCAGGAGCTGCGCAG
      N I G R D R A H D M L R L A D D D E V R
8870 CGTGAAGTTACCGTGCCGAGCCAGAGCCTGCTGGTGAGCAGCGGGAACGCCCGAGCC 8929
      +-----+-----+-----+-----+-----+-----+-----+-----+
      GCGACTTCAAGTGGCACGGCTCGGTCTCGGACGACCACTCGTCGCGCCTTGGGGGCTCGG
      G S F N V T G L W L R S T L L A S R G L
8930 GCACGTGCGTCGCGGCGTCGGTGGTCACTGCTGTTCTCTCTCCCTGCGGCCAGTTCTCTC 8989
      +-----+-----+-----+-----+-----+-----+-----+-----+
      CGTGACGACGCGCGCAGCCACCAAGTAGACCAAGAGAGGAAGGACGCCCGGTCAAGGAG
      R V H T A A D T T M
      <---ORF7 * R P E R R G A A L E E
9049 GCAGATGCCGACGACCTCGGCCCGGTGACGGCTCCGCGAGCATGTCGTCGCGCATCCCGCGC 9049
      +-----+-----+-----+-----+-----+-----+-----+-----+
      CGTCTACGGCTGCTGGAGCCGGCCACTGCCGAGGCGCTCGTACAGCAGCGCGTAGGCGCG
      C I G V V E A P S P E A L M D D R M R A
9109 CGCGCCGCGCGGTGGGCCCGGTGTCGAGGACCCGCTTCACCGACTCCCGGAGCTGGTC 9109
      +-----+-----+-----+-----+-----+-----+-----+-----+
      GCGGCGCGCGCCACCCGGCCAGCAGCTCCTGGGCGAAGTGGCTGAGGGCCTCGACACG
      A G A R H A P D D L V R K V S E R L Q D
9110 GGGGTCAGCTCGGGCACGGGACGCGGATCCCCGCCCCCAATTCTGCGTGGCTGGCGTGGCG 9169
      +-----+-----+-----+-----+-----+-----+-----+-----+
      CCCCCAGTCGAGCCCGTGCCCCGTGCGCTAGGGCGGGGCTTAAGGACGACGCGGACGCG
      P T L E P V P L A I G A G F E Q T R Q A

```

FIGURE 2

5/60

```

9170 GCGACGCCGGTGTCCAGCCGTCGGGCAGGATACCTGGGCACGCCGTGGATCGCCGC 9229
+-----+-----+-----+-----+-----+-----+-----+-----+
CGGTGGGCCACAGGGTCGGCAGCCCGTCTTAGTGACGCCGTGGGCACCTAGCGGCG
R V G T D W G D P L I V Q P V G H I A A

9230 GGTGTGCCAGCTCCCGGGTCCCGCGTGGTGCACCGTCGCCCGCAGGTCCGCGACAGCGC 9289
+-----+-----+-----+-----+-----+-----+-----+-----+
CCACACGGTCGAGGGCCAGCGCGCACACAGTGGCAGCGCGCTCCAGCCGTGTCGCGC
T H W S G P G G H V T A A C T P L L A

9290 GTGCATCGGGACGAAGCCGACCGTCGGGACGTGTCCGGGATGTTCCGCGACGCCTTC TAG 9349
+-----+-----+-----+-----+-----+-----+-----+-----+
CACGTAGCCCTGCTTCGGCTGGCACGCCCTGCAACAGGCCCTACAAAGCGTCCGGAAGATC
H M P V F G V T R V N D P I N A V G E L

9350 CTGCTGCGCGTCGAAGGTGCGGATGATCTCGGCGTCGACGTCGCCGACGGCACCCAGCAG 9409
+-----+-----+-----+-----+-----+-----+-----+-----+
GACGACGCGCAGCTTCAGCGCTACTAGAGCCGCGCAGCTGCAGCGGCTGCCGTGGTCTGTC
Q Q A D F T A I I E A D V D G V A G L L

9410 CTCCTCGATGGAGACCTGCCCGGATGCTGTTCTCGCGGCTGGAGATCCCGAGCGTGAGGCA 9469
+-----+-----+-----+-----+-----+-----+-----+-----+
GAGGAGCTACCTCTGGACGGGTACGACAAGAGCGCCGACCTCTAGGGCTCGCACCTCCGT
E E I S V Q G I S N E R S S I G L T L C

9470 CACGCGGCGGCTCGGGCTCGTGTGAGCCATTCCGGCACCCACGACGGCCCGTTGTA 9529
+-----+-----+-----+-----+-----+-----+-----+-----+
GTGCGCGCGCGAGCCGAGCAGCAGCTCGGTAAGGCCGTGGTGCCCTGCCGGGCAACAT
V R R R E P E D H L W E P V V S P G N Y

9530 GTCGACGTAGCGCATCCCGACGGTCTTCAGGCCGGTGTGAGACCTGATCGCGGCCGGGCGC 9589
+-----+-----+-----+-----+-----+-----+-----+-----+
CAGCTGCATCGGTAGGGCTGCCAGAAGTCCGGCCACAGCTCGGACTAGCGCCGGCCCCG
D V Y R M G V T K L G T D L R I A A P A

```

FIGURE 2

6/60

```

9590  GGGGTCGATCGTCCACTGCCCCGACGACCACTCCCTCGTCGAAGCCGGCCGCGCTACTT      9649
      +-----+-----+-----+-----+-----+-----+-----+-----+
      CCCCAGCTAGCAGGTGACGGGCTGCTGGTGGAGGAGCAGCTTCCGGCCCGCGGCGCATGAA
      P D I T W Q G V V V E E D F A P G G Y K

9650  CTCCAGCGTCCAGGTGAGCCACTCGGGGAGCGGGTCTCCCGGTGCTCCTCCGGCTGGTC      9709
      +-----+-----+-----+-----+-----+-----+-----+-----+
      GAGTCCGAGGTCCACTCGGTGAGCCGCTCGCCCAAGAGGGCCACGAGGAGGCCGACCCAG
      E L T W T L W E A L P D E R H E E P Q D

9710  GGGCAGCAGGCCGAGGAAGTTCTGCCGCGCCCGGGTGGTGATGTCGGGTCCCCACAGCAG      9769
      +-----+-----+-----+-----+-----+-----+-----+-----+
      CCCGTCGTCCGGCTCCTTCAAGACGGCGCGGCCCAACCACTACAGCCCCAGGGGTGTCGTC
      P L L G L F N Q R A R T T I D P G W L L

9770  CCGCGCGTGCGGCGTTCCGGTCACCGCGCGCGGATGGCGCGGCGAAGGTGAGCGGCTC      9829
      +-----+-----+-----+-----+-----+-----+-----+-----+
      GGCGCGCACGCCGCAAGGCCAGTGGCGCGCGGCTACCCGCGCGCTTCCACTCGCCCGAG
      R A H P T G T V A A A I P A A F T L P E

9830  CCAGATGACCAGGTCGGGCGCCCACTTCCGGCAGAACGAGACCATGCTTCGATGAGCGT      9889
      +-----+-----+-----+-----+-----+-----+-----+-----+
      GGTCTACTGGTCCAGCCCGCGGTGAAGCCGCTTGTCTGTGTACGGAAGCTACTCGCA
      W I V L D P R W K R C F S V M G E I L T

9890  GTCCGGGCTCATCAGGGCGTAGAAGGTCTGGGGTGAGCACGGTCTGTCATGCCCGCAGCGTG      9949
      +-----+-----+-----+-----+-----+-----+-----+-----+
      CAGGCCCGAGTAGTCCCGCATCTCCAGCCCCACCTGTGCCAGACGTAACGGGTCTGTCAC
      D P S M L A Y F T P T L V T Q M G L L H

9950  CTCCCAGGTCAAGGTGGGGGTCCCGCTCGCTGAAGTCCAGGCTCCCGACGTAGTCGAT      10009
      +-----+-----+-----+-----+-----+-----+-----+-----+
      GAGGGTCCAGTTCACCGCCCCAGGGCGAGCGACTTCAGGTCCGAGGSCCTGCATCAGCTA
      E W T L T A P D R E S F D L S R V Y D I

```

FIGURE 2

```

10010  GATGTCGTGGCCCCGCGTGGGTTCATGAAGTCCACGAGGTCGACGTCGGTGCCGACCGGGAC 10069
      +-----+-----+-----+-----+-----+-----+-----+-----+
      CTACAGCACCGGGCGCACCCAGTACTTCAGGTGCTCCAGCTGCAGCCACGGCTGGCCCTG
      I D H G A H T M F D V L D V D T G V P V
10070  GCGCGTCAGCCCCGCGCGGTGATGTCTCGGTGAGCGCCGGGGACGCGACCGCGGAC 10129
      +-----+-----+-----+-----+-----+-----+-----+-----+
      CCGCCAGTCGGGGCGCGCCACTACAGGAGCCACTCGCGGCCCTGCGCTGGTGGCCTG
      A T L G A A T I D E T L A P S A V R V
10130  CTCGTGCCCCCGCGCGGAACGCCCATGCGAGGGGACGAGGCCGAAGAGGTGGCTCTT 10189
      +-----+-----+-----+-----+-----+-----+-----+-----+
      GAGCACGGGGCGCGCCTTGCGGGTACGCTCCCCCTGCTCCGGCTTCTCCACCGAGAA
      E H G A A R F A W A L P V L G F L H S K
10190  GCTGGCCATGAGGAGAGACGACGCGCATCGCGGTTACCTCAGAGCTCGACGGGCAGC 10249
      +-----+-----+-----+-----+-----+-----+-----+-----+
      CGACCGGTACCTCCTCTTCTGTGCGCGTAGCGCCAATGGAGTCTCGAGCTGCCCCGTCG
      S A M S S F V V R M * L E V P C
      <---ORF8
10250  GGTGTGTTCCCGCAGGACGGGTGATCGCGCGCGCCGACACCGGGCCGCTGGGCGTGA 10309
      +-----+-----+-----+-----+-----+-----+-----+-----+
      CCAACCAAGGGCGTCTTGGCCACTAGCCGCGCGGCTGCTGGCCCCGGCGACCCGCACT
      R N T G R L V P S R R R R V V P G S P T
10310  GTCCGGCAGCGCCTTGCGCGCGCGCCGCGAGTGCGGGGTGCGAGCGGGTGACCACT 10369
      +-----+-----+-----+-----+-----+-----+-----+-----+
      CAGCCCCGTGCGGGAACCGGCGCGCGGGCGTACGCCCCCACCCTCGCGCCCACTGGTCGA
      L G P L A K A A A R L A A T A L A T V L
10370  CCTCCAGCCTGCCGGGGTGCGCGCGATGTCCGACAGCGCGGGTGCGGCGTGGGGCGGT 10429
      +-----+-----+-----+-----+-----+-----+-----+-----+
      GGAGGTCGACGGCCCCACCGCGCGTACAGGCTGTGCGGCGGCCAGCGCAGCCCCGCCA
      E E L R G P H G R H A S L A R D A D P R

```

FIGURE 2

8/60

10430 CCACGTCGAGGCGGTCGGGCTCGGCGAAGACCTCCGGGTGCGGGTTGGCCGCCGCGACGA 10489
 +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 GGTGCAGCTCCGCCAGCCCGAGCCGCTTCTGGAGGCCCGAGCCCAACCGCGGCGCTGCT
 D V D L R D P E A F V E P D R N A A V

 10490 CGACCACGACCTCCTCGCCTTCGCCGATCACGTGCTGCCCGAGCCGACCTCTCGGGTGG 10549
 +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 GCTGGTCTGGAGGAGCGGAAGCGGCTAGTGCACGAGCGGCTCGGCGTGGAGACGCCACC
 V V V V E E G E G I V H E G L R V E A T

 10550 CCGTGCGCCGCTCCAGGTGCAATGCCGGGTGCAGGCGCAGCACCTCGGCGACGGTTCGCT 10609
 +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 GGCACGCGGCGAGGTCCACGTTACGGCCACGTCCGCTCGTGGAGCCGCTGCCAAGCGA
 A T R R E L H L A P H L R L V E A V T R

 10610 GCGCGGGCGGGGTCGTCGGCGATCCGTTTCGGCCAGCCCCGGTTCGGCCGAGACGGCCA 10669
 +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 CGCGCCGCCGCCAGCAGCCGCTAGSCAAGCCGGTTCGGGCCAAGCCGGCTCTGCCGGT
 Q A A A P D D A I R E A L G P E A S V A

 10670 GGACCGGTCGACCACGGTGTTCGGGGTCACTCTGGCCCCGGCGAACAGGGCGCGCAGTG 10729
 +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 CCTGGCGCAGCTGGTCCACAAAGCGCCAGTAGAGCCGGGGCCGCTTGTCGCCGCGTCAC
 L V A D V V T N A T M E A G A F L A R L

 10730 CGGGTCCGGCGGCGAGTGCCGCGACCGCTGCTTCGGTCAACCGGAGCTGTCGGGGCTGA 10789
 +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 GCCCAGCCGCCGTCACGGCGTGGCGACGAAGCCAGTGGCGCTCGACGACGCCCGACT
 A P D A P L A A V A A E T V A L Q Q P S

 10790 GCTGGGGTCCAGGTGACCGGGCGTCCCACGCGCGCCCGCGCAGCAC'TCCGGCTGCGC 10849
 +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 CGACCCGCGAGTCCGACTGGGCCCGCAGGGTGGCCGCGGCGGCTGCTGAGGCCGACGCG
 L Q A D L S V R A D W A A G R L V G A A

FIGURE 2

9/60

10850 CGAGCAGCGCGGTATGCCCTGCACCGGTACCTGCCAGGCGAAGTCGCCGACCGAGGTCCA 10909
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
GCTCGTCCGCCAGTACGGACGTGGCCATGGACGGTCCGCTTCAGCGGCTGGTCCAGGT
G L V A T M G Q V P V Q W A F D G V L D
10910 GCCGCGCGCCCGCGGAGCAGACCGGCGAAGCTCTCCGCCAGTTCCTCCGACGTCCG 10969
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
CGCGCGCGCGCGCGCCCTCGTCTGGCCGCTTCGAGAGCGGTCAAGGGCTGCAGCC
L R A G A G P L L G A F S E A L E G V D
10970 GGACCTCGCCTTCCCAGGACGCGCGGTGCACGTCCCGGAACGGCTGGGCCCACTCGGCGG 11029
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
CCTGGAGCGGAAGGTCTCTGCGCCGACGTGCAGGGCTTGCCGACCCCGGTGAGCCGCC
P V E G E W S A A H V D R F P Q A W E A
11030 GTGGCGCGCGCGCGCATCCATTCCGGTGTCCGTCGCGGTGGCGCGGTGAACGCGG 11089
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
CACCGCGCGCGCGCGGTAGGTAAGCCACACGACGAGGCCACCGGCCCACTTGCGCC
P P A G A A R M W E P T R G T A R T F A
11090 GGTCGTCGAGACCTGCCGCGCGGTGGCGTGGTCCGCCACCACCCACGTCCTCGGTGCGGC 11149
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
CCAGCAGCTCGTGGACGCGCGCGCCACCGCACCGCGGTGGTGGTGCAGAGCCACGCGG
P D D L V Q R A T A H D A V V W T E T R
11150 TGC CGCGCACACCGGACTCGCGCATCGAGCGGTACCGCGCTGCGGGTCGTCTGTGTC 11209
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
ACGCGCGGTGTGGCCTGAGCGGTAGTCCGCCATGGCCGCGACGCCAGCAGCAGCACAG
S R R V G S E R M S R Y R Q P D D H
CGCACAGCAGCATCGGTAAGGTGCGCGTGTGCGGTAAACCCAGTGCAGGCGCGGA
11210 +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 11269
GCGTGTGTCGTAGCCCATTCACAGCGGCAACGAGCGCATTTGGGTACGTCCGGCGCCT
G C L L M P Y P D G N S G Y G W H L G R

FIGURE 2

10/60

```
11270 TCATCTGGAGCTGCCCTGCCAGCCCGCGCGGATCGGTCGTGGTCATGAATTC 11321
      +-----+-----+-----+-----+-----+-----+
      AGTAGACCTCGACGGACGGGTCGGGCGCGCGCTAGCCAGCACCAGTACTTAAG
      I M Q L Q R G L G A R D T T T M
      <-----ORF9
      * S N
      <-----ORF10 ...
```

FIGURE 2

11/60

```

43610      43630      43650
TTTGACAGGTCCGCCACGCGTCCCCCTACTCGACGACCACGCAATGGGCGAACAATATAG
43670      43690      43710
GAAGGATCAAGAGGTTGACATCGCCTCGTCGAGCCAACGAACCTGTGAACATCTGCATGT
43730      43750      43770
TGACAAGATCAACGGCGGCTACCTACTGTGGTGGCCCAGTGACGGGTTGCCGCACATCGC
43790      43810      43830
GCTGGGGAGATTCTTTGAATTTGCGCCGTAGCACCGACCTGGAAAGCGAGCAAATGCTCC
43850      43870      43890
GGTGAATGGGATCAGTGATTCCCCGCGTCAATTGATCACCTTCTGGGCGCTTCCGGCTT
V N G I S D S P R Q L I T L L G A S G F
ORF13 ---->
43910      43930      43950
CGTCGGGAGCGCGGTTCTGCGCGAGCTGCGCGACCACCCGGTCCGGCTGCGCGCGGTGTC
V G S A V L R E L R D H P V R L R A V S
43970      43990      44010
CCGCGGCGGAGCGCCCGCGGTTCCGCCCCGGCGCCGCGGAGGTCGAGGACCTGCGCGCCGA
R G G A P A V P P G A A E V E D L R A D
44030      44050      44070
CCTGCTGGAACCGGGCCGGGCGCCGCCGCGATCGAGGACGCCGACGTGATCGTGCACCT
L L E P G R A A A A I E D A D V I V H L
44090      44110      44130
GGTGGCGCAGCAGCGGGCGGTTCCACCTGGCGCAGCGCCACCTCCGACCCGGAAGCCGA
V A H A A G G S T W R S A T S D P E A E
44150      44170      44190
GCGGGTCAACGTCGGCCTGATGCACGACCTCGTCGGCGCGCTGCACGATCGCCGCGAGTTC
R V N V G L M H D L V G A L H D R R R S
44210      44230      44250
GACGCCGCCCGTGTTGCTCTACGCGAGCACCGCACAGGCCGCGAACCCGTCGGCGGCCAG
T P P V L L Y A S T A Q A A N P S A A S
44270      44290      44310
CAGGTACGCGCAGCAGAAGACCGAGGCCGAGCGCATCCTGCGCAAAGCCACCGACGAGGG
R Y A Q Q K T E A E R I L R K A T D E G
44330      44350      44370
CCGGGTGCGCGGCGTGATCCTGCGGCTGCCCGCCGTCTACGGCCAGAGCGGCCCGTCCGG
R V R G V I L R L P A V Y G Q S G P S G
44390      44410      44430
CCCCATGGGGCGGGCGTGCTGCGCAGCGATGATCCGGCGTGCCCTCGCCGGCGAGCCGCT
P M G R G V V A A M I R R A L A G E P L
44450      44470      44490
CACCATGTGGCACGACGGCGGCGTGCGCCGCGACCTGCTGCACGTGAGGACGTGGCCAC
T M W H D G G V R R D L L H V E D V A T
44510      44530      44550
CGCGTTCGCCCGCGCTGGAGCACCACGACGCGCTGGCCGGCGGCACGTGGGCGCTGGG
A F A A A L E H D A L A G G T W A L G
44570      44590      44610
CGCCGACCGATCCGAGCCGCTCGGCGACATCTTCCGGGCCGTCTCCGGCAGCGTCGCCCCG
A D R S E P L G D I F R A V S G S V A R
44630      44650      44670
GCAGACCGGCAGCCCCGCGCTCGACGTGGTCACCGTGCCCGCGCCCCGAGCACGCCGAGGC
Q T G S P A V D V V T V P A P E H A E A

```

FIGURE 3

12/60

44690 44710 44730
CAACGACTTCCGCAGCGACGACATCGACTCCACCGAGTTCCGCAGCCGGACCGGCTGGCG
N D F R S D D I D S T E F R S R T G W R
44750 44770 44790
CCCCCGGGTTTCCCTCACCGACGGCATCGACCGGACGGTGGCCGCCCTGACCCCCACCGA
P R V S L T D G I D R T V A A L T P T E
44810 44830 44850
GGAGCACTAGTGCGGGTACTGCTGACGTCCTTCGCGCACCGCACGCACTTCCAGGGACTG
E H *
V R V L L T S F A H R T H F Q G L
ORF14 ---->
44870 44890 44910
GTCCCGCTGGCGTGGGCGCTGCGCACCGCGGGTCACGACGTGCGCGTGGCCGCCAGCCC
V P L A W A L R T A G H D V R V A A Q P
44930 44950 44970
GCGCTACCGACGCGGTTCATCGGCGCCGGTCTACCGCGGTACCGCTCGGCTCCGACCAC
A L T D A V I G A G L T A V P V G S D H
44990 45010 45030
CGGCTGTTTCGACATCGTCCCGGAAGTCGCCGCTCAGGTGCACCGCTACTCCTTCTACCTG
R L F D I V P E V A A Q V H R Y S F Y L
45050 45070 45090
GACTTCTACCACCGCGAGCAGGAGCTGCACTCGTGGGAGTTCCTGCTCGGCATGCAGGAG
D F Y H R E Q E L H S W E F L L G M Q E
45110 45130 45150
GCCACCTCGCGGTGGGTATACCCGGTGGTCAACAACGACTCCTTCGTCGCCGAGCTGGTC
A T S R W V Y P V V N N D S F V A E L V
45170 45190 45210
GACTTCGCCCCGGGACTGGCGTCTGACCTGGTGTCTGCGGAGCCGTTACCTTCGCCGGC
D F A R D W R P D L V L W E P F T F A G
45230 45250 45270
GCCGTCGCGGCCCGGGCCTGCGGAGCCGCGCACGCCCGGCTGCTGTGGGGCAGCGACCTC
A V A A R A C G A A H A R L L W G S D L
45290 45310 45330
ACCGGCTACTTCCGCGGCCGGTTCCAGGCGCAACGCCTGCGACGGCCCGGAGGACCGG
T G Y F R G R F Q A Q R L R R P P E D R
45350 45370 45390
CCGGACCCGCTGGGCACGTGGCTGACCGAGGTGCGGGGCGCTTCGGCGTCAATTCGGC
P D P L G T W L T E V A G R F G V E F G
45410 45430 45450
GAGGACCTCGCGGTGCGGCAGTGGTGGTGCACGAGTTGCCGCGGAGTTTCCGGCTGGAC
E D L A V G Q W S V D Q L P P S F R L D
45470 45490 45510
ACCGGAATGGAAACCGTTGTGCGCGGACCCCTGCCCTACAACGGCGCGTGGTGGTTCCG
T G M E T V V A R T L P Y N G A S V V P
45530 45550 45570
GACTGGCTCAAGAAGGGCAGTGCAGCTCGACGCATCTGCATTACCGGAGGGTTCTCCGGA
D W L K K G S A T R R I C I T G G F S G
45590 45610 45630
CTCGGGCTCGCCGCCGATGCCGATCAGTTGCGCGGACGCTCGCGCAGCTCGCGCGATTC
L G L A A D A D Q F A R T L A Q L A R F
45650 45670 45690
GATGGCGAAATCGTGGTTACGGGTTCGGGTCCGGATACCTCCGCGGTACCGGACAACATT
D G E I V V T G S G P D T S A V P D N I

FIGURE 3

13/60

```

45710      45730      45750
CGTTTGGTGGATTTCGTTCCGATGGGCGTTCTGCTCCAGAACTGCGCGGCGATCATCCAC
R L V D F V P M G V L L Q N C A A I I H
45770      45790      45810
CACGGCGGGCCGGAACCTGGGCCACGGCACTGCACCACGGAATTCCGCAAATATCAGTT
H G G A G T W A T A L H H G I P Q I S V
45830      45850      45870
GCACATGAATGGGATTGCATGCTACGCGGCCAGCAGACCGCGGAATGGGCGCGGGAATC
A H E W D C M L R G Q Q T A E L G A G I
45890      45910      45930
TACCTCCGGCCGGACGAGGTCGATGCCGACTCATTGGCGAGCGCCCTCACCCAGGTGGTC
Y L R P D E V D A D S L A S A L T Q V V
45950      45970      45990
GAGGACCCACCTACACCGAGAACGCGGTGAAGCTTCGCGAGGAGGCGCTGTCCGACCCG
E D P T Y T E N A V K L R E E A L S D P
46010      46030      46050
ACGCCGCGAGGAGATCGTCCCGCGACTGGAGGAATCACGCGCCGCCCGCTAGCGG
T P Q E I V P R L E E L T R R H A G *
46070      46090      46110
TTTCCGACCGACAAGTCCGTCCGACAGCACACCTCCGGAGGGAGCAGGGATGTACGAGGG
M Y E G
ORF15 ---->
46130      46150      46170
CGGGTTCGCGGAGCTTTACGACCGGTTCTACCGCGGCGCGGGGCAAGGACTACGCGGCCGA
G F A E L Y D R F Y R G R G K D Y A A E
46190      46210      46230
GGCCGCGCAGGTCGCGCGGCTGGTCAGAGACCGCCTGCCCTCGGCTTCCTCGCTGCTCGA
A A Q V A R L V R D R L P S A S S L L D
46250      46270      46290
CGTGGCCTGCGGGACCGGCACCCACCTGCGCCGGTTGCGCGACCTCTTCGACGACGTGAC
V A C G T G T H L R R F A D L F D D V T
46310      46330      46350
CGGGCTGGAGCTGTGCGCGGCGATGATCGAGGTGCGCCGCGCGCAGCTCGGCGGCATCCC
G L E L S A A M I E V A R P Q L G G I P
46370      46390      46410
GGTGCTGCAGGGCGACATGCGCGACTTCGCGCTGGATCGCGAGTTCGACGCCGTCACCTG
V L Q G D M R D F A L D R E F D A V T C
46430      46450      46470
CATGTTTCAGCTCCATCGGGCACATGCGCGACGGCGCCGAGCTGGACCAGGCGCTGGCGTC
M F S S I G H M R D G A E L D Q A L A S
46490      46510      46530
CTTCGCCCCGCCACCTCGCCCCCGGCGCGTCTGTTGGTGGTTCGAACCGTGGTGGTTCCCGGA
F A R H L A P G G V V V V E P W W F P E
46550      46570      46590
GGAATTCTCGACGGCTACGTGGCCGGTGACGTGGTGGCGGACGGCGACCTGACGATCTC
D F L D G Y V A G D V V R D G D L T I S
46610      46630      46650
GCGCGTCTCGCACTCCGTGCGCGCCCGGCGCGACCCGGATGGAGATCCACTGGGTGCTG
R V S H S V R A G G A T R M E I H W V V
46670      46690      46710
GGCCGACGCGGTGAACGGTCCGCGGCACACGTGGAGCACTACGAGATCACGCTCTTCGA
A D A V N G P R H H V E H Y E I T L F E

```

FIGURE 3

14/60

46730 46750 46770
GCGGCAGCAGTACGAGAAGGCCCTTCACCGCGGGCGGTTGCGCTGTGCAGTACCTGGAGGG
R Q Q Y E K A F T A A G C A V Q Y L E G
46790 46810 46830
CGGACCTCCGGACGCGGGTTGTTTCGTTCGGTGTGCGCGGATGACCCGTGCGTTTCGCGTTT
G P S G R G L F V G V R G *
46850 46870 46890
TCCGTTCTTGGCACAGGTGATCCGCTCCACGGGCCCCCTTCCCCGCCGTGACCGGACCCTT
46910 46930 46950
ACAGTGAGTGCGGGTCTTGATCGACAACGCCCGGCGGCAGCAAGCGGAGCCGTCGACGAC
V R V L I D N A R R Q Q A E P S T T
ORF16 --->
46970 46990 47010
ACCGCAGGGAGAGTCGATGGGTGATCGGACCGGCGACCGGACGATTCCGGAATCCTCGCA
P Q G E S M G D R T G D R T I P E S S Q
47030 47050 47070
GACCGCAACGCGTTTCTGCTCGGCGACGGCGGAATCCCCACCGCCACGGCGGAAACCCA
T A T R F L L G D G G I P T A T A E T H
47090 47110 47130
CGACTGGCTGACCCGCAACGGCGCCGAGCAGCGGCTCGAGGTGGCGCGCGTGCCTTCAG
D W L T R N G A E Q R L E V A R V P F S
47150 47170 47190
CGCCATGGACCGCTGGTCGTTCCAGCCCGAGGACGGCAGGCTCGCCACGAGTCCGGGCG
A M D R W S F Q P E D G R L A H E S G R
47210 47230 47250
CTTCTTCTCCATCGAGGGCCTGCACGTGCGGACGAACCTTCGGCTGGCGGGCGGACTGGAT
F F S I E G L H V R T N F G W R R D W I
47270 47290 47310
CCAGCCCATCATCGTGCAGCCCGAGATCGGCTTCCTCGGCCTCATCGTCAAGGAGTTCGA
Q P I I V Q P E I G F L G L I V K E F D
47330 47350 47370
CGGTGTGCTGCACGTGCTGGCGCAGGCCAAGGCCGAGCCGGGCAACATCAACGCCGTCCA
G V L H V L A Q A K A E P G N I N A V Q
47390 47410 47430
GCTCTCCCCGACCCTGCAGGCGACCCGAGCAACTACACCGCGCTCCACCGCGCTCGAA
L S P T L Q A T R S N Y T G V H R G S K
47450 47470 47490
GGTCCGGTTTCATCGAGTACTTCAACGGCAGCGCCCCGAGCCGGATCCTCGTTCGACGTGCT
V R F I E Y F N G T R P S R I L V D V L
47510 47530 47550
CCAGTCCGAGCAGGGCGCGTGGTTCTGCGCAAGCGCAACCGGAACATGGTTCGTCGAGGT
Q S E Q G A W F L R K R N R N M V V E V
47570 47590 47610
GTTTCGACGACCTGCCCCGAGCACCCGAACCTCCGGTGGCTGACCGTCGCGCAGCTGCGGGC
F D D L P E H P N F R W L T V A Q L R A
47630 47650 47670
GATGCTGCACCACGACAACGTGGTGAACATGGACCTGCGCACCGTGTGTCCTGCGTCCC
M L H H D N V V N M D L R T V L A C V P
47690 47710 47730
GACCGCCGTGGAGCGGGACCGGGCCGACGACGTGCTCGCGCGCCTGCCCCAGGGCTCGTT
T A V E R D R A D D V L A R L P E G S F

FIGURE 3

15/60

47750 47770 47790
 CCAGGCCCCGGCTGCTGCACTCGTTCATCGGCGCGGGCACCCCGGCCAACACATGAACAG
 Q A R L L H S F I G A G T P A N N M N S
 47810 47830 47850
 CCTGCTGAGCTGGATCTCCGACGTGCGCGCCAGGCGCGAGTTCGTGCAGCGCGCCGCC
 L L S W I S D V R A R R E F V Q R G R P
 47870 47890 47910
 GCTGCCCCGACATCGAGCGCAGCGGGTGGATCCGCCGCGACGACGGCATCGAGCACGAGGA
 L P D I E R S G W I R R D D G I E H E E
 47930 47950 47970
 GAAGAAGTACTTCGACGTCTTCGGCGTACACGGTGGCGACCAGCGACCGCGAGGTCAACTC
 K K Y F D V F G V T V A T S D R E V N S
 47990 48010 48030
 GTGGATGCAGCCGCTGCTCTCGCCCGCCAACAACGGCCTGCTCGCCCTGCTGGTCAAGGA
 W M Q P L L S P A N N G L L A L L V K D
 48050 48070 48090
 CATCGGCGGCACGTTGCACGCGCTCGTGCAGCTGCGCACCGAGGCGGGCGGGATGGACGT
 I G G T L H A L V Q L R T E A G G M D V
 48110 48130 48150
 CGCCGAGCTGGCGCCTACGGTGCAGTCCAGCCCCGACAACTACGCCGACGCGCCCGAGGA
 A E L A P T V H C Q P D N Y A D A P E E
 48170 48190 48210
 GTTCCGACCGGCCTATGTGGACTACGTGTTGAACGTGCCGCGCTCGCAGGTCCGCTACGA
 F R P A Y V D Y V L N V P R S Q V R Y D
 48230 48250 48270
 CGCATGGCACTCCGAGGAGGGCGCCGGTTCTACCGCAACGAGAACC GGTTACATGCTGAT
 A W H S E E G G R F Y R N E N R Y M L I
 48290 48310 48330
 CGAGGTGCCCCGCCACTTCGACGCCAGTGCCGCTCCCGACCACCGGTGGATGACCTTCGA
 E V P A D F D A S A A P D H R W M T F D
 48350 48370 48390
 CCAGATCACCTACCTGCTCGGGCACAGCCACTACGTCAACATCCAGCTGCGCAGCATCAT
 Q I T Y L L G H S H Y V N I Q L R S I I
 48410 48430 48450
 CGCGTGCGCCTCGGCCGTCTACACCAGGACCGCCGGATGAAACGCGCGCTGACCGACCTG
 A C A S A V Y T R T A G *
 M K R A L T D L
 ORF17 --->
 48470 48490 48510
 GCGATCTTCGGCGGCCCGAGGCATTCCTGCACACCCTCTACGTGGGCAGGCCGACCGTC
 A I F G G P E A F L H T L Y V G R P T V
 48530 48550 48570
 GGGGACCGGGAGCGGTTCCTTCGCCCCGCTGGAGTGGGCGCTGAACAACAACCTGGCTGACC
 G D R E R F F A R L E W A L N N N W L T
 48590 48610 48630
 AACGGCGGACCACTGGTGC GCGAGTTCGAGGGCCGGGTGCGCGACCTGGCGGGTGTCCGC
 N G G P L V R E F E G R V A D L A G V R
 48650 48670 48690
 CACTGCGTGGCCACCTGCAACGCGACGGTTCGCGCTGCAACTGGTGCTGCGCGCGAGCGAC
 H C V A T C N A T V A L Q L V L R A S D
 48710 48730 48750
 GTGTCCGGCGAGGTGCTCATGCCTTCGATGACGTTTCGCGGCCACCGCGCACGCGCGAGC
 V S G E V V M P S M T F A A T A H A A S

FIGURE 3

16/60

48770 48790 48810
TGGCTGGGGCTGGAACCGGTGTTCTGCGACGTGGACCCCGAGACCGGCTGCTCGACCCC
W L G L E P V F C D V D P E T G L L D P
48830 48850 48870
GAGCACGTCGCGTCGCTGGTGACACCGCGGACGGGCGCGATCATCGGCGTGACCTGTGG
E H V A S L V T P R T G A I I G V H L W
48890 48910 48930
GGCAGGCCCCGCTCCGGTTCGAGGCGCTGGAGAAGATCGCCGCCGAGCACCAGGTCAAACCTC
G R P A P V E A L E K I A A E H Q V K L
48950 48970 48990
TTCTTCGACGCCGCGCACGCGCTGGGGCTGCACCGCCGGCGGGCGGCCGGTTCGGCGCCTTC
F F D A A H A L G C T A G G R P V G A F
49010 49030 49050
GGCAACGCGCGAGGTGTTTCAGCTTCCACGCCACGAAGGCGGTACCTCGTTCGAGGGCGGC
G N A E V F S F H A T K A V T S F E G G
49070 49090 49110
GCCATCGTCACCGACGACGGGCTGCTGGCCGACCGCATCCGCGCCATGCACAACCTTCGGG
A I V T D D G L L A D R I R A M H N F G
49130 49150 49170
ATCGCACCGGACAAGCTGGTGACCGATGTGCGGCACCAACGGCAAGATGAGCGAGTGCGCC
I A P D K L V T D V G T N G K M S E C A
49190 49210 49230
GCGGCGATGGGCCTCACCTCGCTCGACGCCTTCGCCGAGACCAGGGTGCACAACCGCCTC
A A M G L T S L D A F A E T R V H N R L
49250 49270 49290
AACCACGCGCTCTACTCCGACGAGCTCCGCGACGTGCGCGGCATATCCGTGCACGCGTTC
N H A L Y S D E L R D V R G I S V H A F
49310 49330 49350
GATCCTGGCGAGCAGAACAACCTACCAGTACGTGATCATCTCGGTGGACTCCGCGGCCACC
D P G E Q N N Y Q Y V I I S V D S A A T
49370 49390 49410
GGCATCGACCGCGACCAAGTTGCAGGCGATCCTGCGAGCGGAGAAGGTTGTGGCACAACCC
G I D R D Q L Q A I L R A E K V V A Q P
49430 49450 49470
TACTTCTCCCCGGGTGCCACCAGATGCAGCCGTACCGGACCGAGCCCGCTGCGGCTG
Y F S P G C H Q M Q P Y R T E P P L R L
49490 49510 49530
GAGAACACCGAACAGCTCTCCGACCGGGTGCTCGCGCTGCCCACCGGCCCCGCGGTGTCC
E N T E Q L S D R V L A L P T G P A V S
49550 49570 49590
AGCGAGGACATCCGGCGGGTGTGCGACATCATCCGGCTCGCCGCCACCAGCGGCGAGCTG
S E D I R R V C D I I R L A A T S G E L
49610 49630 49650
ATCAACGCGCAATGGGACCAGAGGACGCGCAACGGTTCGTGACGACCTGCGCCACAAGTG
I N A Q W D Q R T R N G S *
49670 49690 49710
CCAGGAGGTTTCGCTCCCCGATGAACACAACCTCGTACGGCAACCGCCAGGAAGCGGGGT
M N T T R T A T A Q E A G V
ORF18 --->
49730 49750 49770
CGCCGACGCGGCGCGCCCGGACGTGACCGGCGGGCGGTTCGTGCGGGCGCTGAGCTCGGA
A D A A R P D V D R R A V V R A L S S E

FIGURE 3

17/60

```

49790      49810      49830
GGTCTCCCGCGTCAACGGCGCCGGTGACGGTGACGCCGACGTGCAGGCCGCCCGGCTCGC
V S R V T G A G D G D A D V Q A A R L A
49850      49870      49890
CGACCTCGCCGCGCACTACGGGGCGCACCCGTTACGCCGCTGGAGCAGACGCGTGCGCG
D L A A H Y G A H P F T P L E Q T R A R
49910      49930      49950
GCTCGGCCTGGACCGCGCGGAGTTCGCCACCTGCTCGACCTGTTTCGGCCGCATCCCCGA
L G L D R A E F A H L L D L F G R I P D
49970      49990      50010
CCTGGGCACCGCGGTGGAGCACGGTCCGGCGGGCAAGTACTGGTCCAACACGATCAAGCC
L G T A V E H G P A G K Y W S N T I K P
50030      50050      50070
GCTGGACGCCGCGAGGCGCACTGGACGCGGGCGGTCTACCGCAAGCCTGCCTTCCCCTACAG
L D A A G A L D A A V Y R K P A F P Y S
50090      50110      50130
CGTCGGCCTGTACCCCGGGCCGACGTGCATGTTCCGCTGCCACTTCTGCGTGCGGGTGAC
V G L Y P G P T C M F R C H F C V R V T
50150      50170      50190
CGGTGCCCGCTACGAGGCCGCATCGGTCCCGGGCGGCAACGAGACGCTGGCCCGGATCAT
G A R Y E A A S V P A G N E T L A A I I
50210      50230      50250
CGACGAGGTGCCCCACGGACAACCCGAAGGCGATGTACATGTCGGGCGGGCTCGAGCCGCT
D E V P T D N P K A M Y M S G G L E P L
50270      50290      50310
GACCAACCCCGGTCTCGGCGAGCTGGTGTCGCACGCCCGCCGGGCGCGGTTTCGACCTCAC
T N P G L G E L V S H A A G R G F D L T
50330      50350      50370
CGTCTACACCAACGCCTTCGCCCTCACCGAGCAGACGCTGAACCGCCAGCCCGCCTGTG
V Y T N A F A L T E Q T L N R Q P G L W
50390      50410      50430
GGAGCTGGGCGCGATCCGCACGTCCCTCTACGGGCTGAACAACGACGAGTACGAGACGAC
E L G A I R T S L Y G L N N D E Y E T T
50450      50470      50490
CACCGGCAAGCGCGGCGCTTTTGAACGCGTCAAGAAGAACCTGCAGGGCTTCCTGCGGAT
T G K R G A F E R V K K N L Q G F L R M
50510      50530      50550
GCGCGCCGAGCGGGACGCGCCGATCCGGCTCGGCTTCAACCACATCATCCTGCCGGGACG
R A E R D A P I R L G F N H I I L P G R
50570      50590      50610
GGCCGACCGGCTCACCGACCTCGTCGACTTCATCGCCGAGCTCAACGAGTCCAGCCCGCA
A D R L T D L V D F I A E L N E S S P Q
50630      50650      50670
ACGGCCGCTGGACTTCGTGACGGTGCGCGAGGACTACAGCGGCCGCGACGACGGCCGGCT
R P L D F V T V R E D Y S G R D D G R L
50690      50710      50730
GTCGGA TCCGAGCGCAACGAGCTGCGCGAGGGCCTGGTGCGGTTCGTGACTACGCCGC
S D S E R N E L R E G L V R F V D Y A A
50750      50770      50790
CGAGCGGACCCCGGGCATGCACATCGACCTGGGCTACGCCCTGGAGAGCCTGCGGCGGGG
E R T P G M H I D L G Y A L E S L R R G

```

FIGURE 3

18/60

50810 50830 50850
TGTGGACGCCGAGCTGCTGCGCATCCGGCCGAGACGATGCGTCCCACCGCGCACCCCCA
V D A E L L R I R P E T M R P T A H P Q
50870 50890 50910
GGTCGCGGTGCAGATCGACCTGCTCGGCGACGTCTACCTCTACCGCGAGGCGGGCTTCCC
V A V Q I D L L G D V Y L Y R E A G F P
50930 50950 50970
GGAGCTGGAGGGCGCCACCCGCTACATCGCGGGCCGGGTACCCCGTCGACCAGCCTGCG
E L E G A T R Y I A G R V T P S T S L R
50990 51010 51030
CGAGGTGGTGGAGAACTTCGTGCTGGAGAACGAGGGCGTGCAGCCCCGCCCGGCGACGA
E V V E N F V L E N E G V Q P R P G D E
51050 51070 51090
GTACTTCCTCGACGGCTTCGACCAGTCGGTGACCGCACGGCTCAACCAGCTCGAACGAGA
Y F L D G F D Q S V T A R L N Q L E R D
51110 51130 51150
CATCGCCGACGGGTGGGAGGACCACCGCGGCTTCCTGCGCGGAAGGTGAACCGGAGTTGC
I A D G W E D H R G F L R G R *
51170 51190 51210
GAGTACGTGAGCTGGCGGTGGCGGGCGGTTTCGAGTTCACCCCGACCCGAAGCAGGACC
V A G G F E F T P D P K Q D R
ORF19 --->
51230 51250 51270
GGCGGGGCGCTGTTCGTGTCTCCGCTGCAGGACGAGGCGTTCGTGGGCGCGGTGGGCCATC
R G L F V S P L Q D E A F V G A V G H R
51290 51310 51330
GGTTCCCCGTCGCCCAGATGAACCACATCGTCTCCGCCCCGGGCGTGTGCGCGGGCTGC
F P V A Q M N H I V S A R G V L R G L H
51350 51370 51390
ACTTCACCACCACCCCGCCGGGGCAGTGCAAGTACGTCTACTGCGCGCGCGGCCGGGCGC
F T T T P P G Q C K Y V Y C A R G R A L
51410 51430 51450
TCGACGTATCGTCGACATCCGGGTGCGGCTCGCCGACGTTTCGGGAAGTGGGACGCGGTGG
D V I V D I R V G S P T F G K W D A V E
51470 51490 51510
AGATGGACACCGAGCACTTCGGGGCGGTCTACTTCCCCAGGGGCACCGCGCACGCCTTCC
M D T E H F R A V Y F P R G T A H A F L
51530 51550 51570
TCGCGCTTGAGGACGACACCCTGATGTCTGCTACCTGGTCAGCACGCCGTACGTGGCCGAGT
A L E D D T L M S Y L V S T P Y V A E Y
51590 51610 51630
ACGAGCAGGCGATCGACCCGTTTCGACCCCGCGCTGGGTCTGCCGTGGCCCGCGGACCTGG
E Q A I D P F D P A L G L P W P A D L E
51650 51670 51690
AGGTCGTGCTCTCCGACCGCGACACGGTGGCCGTGGACCTGGAGACCGCCAGGCGGCGAG
V V L S D R D T V A V D L E T A R R R G
51710 51730 51750
GGATGCTGCCCGACTACGCCGACTGCGCTCGGCGAGGAGCCCCGCCAGCACCGGCAGGTGAC
M L P D Y A D C L G E E P A S T G R *
* A S Q R P S S G A L V P L H
<--- ORF20...

FIGURE 3

19/60

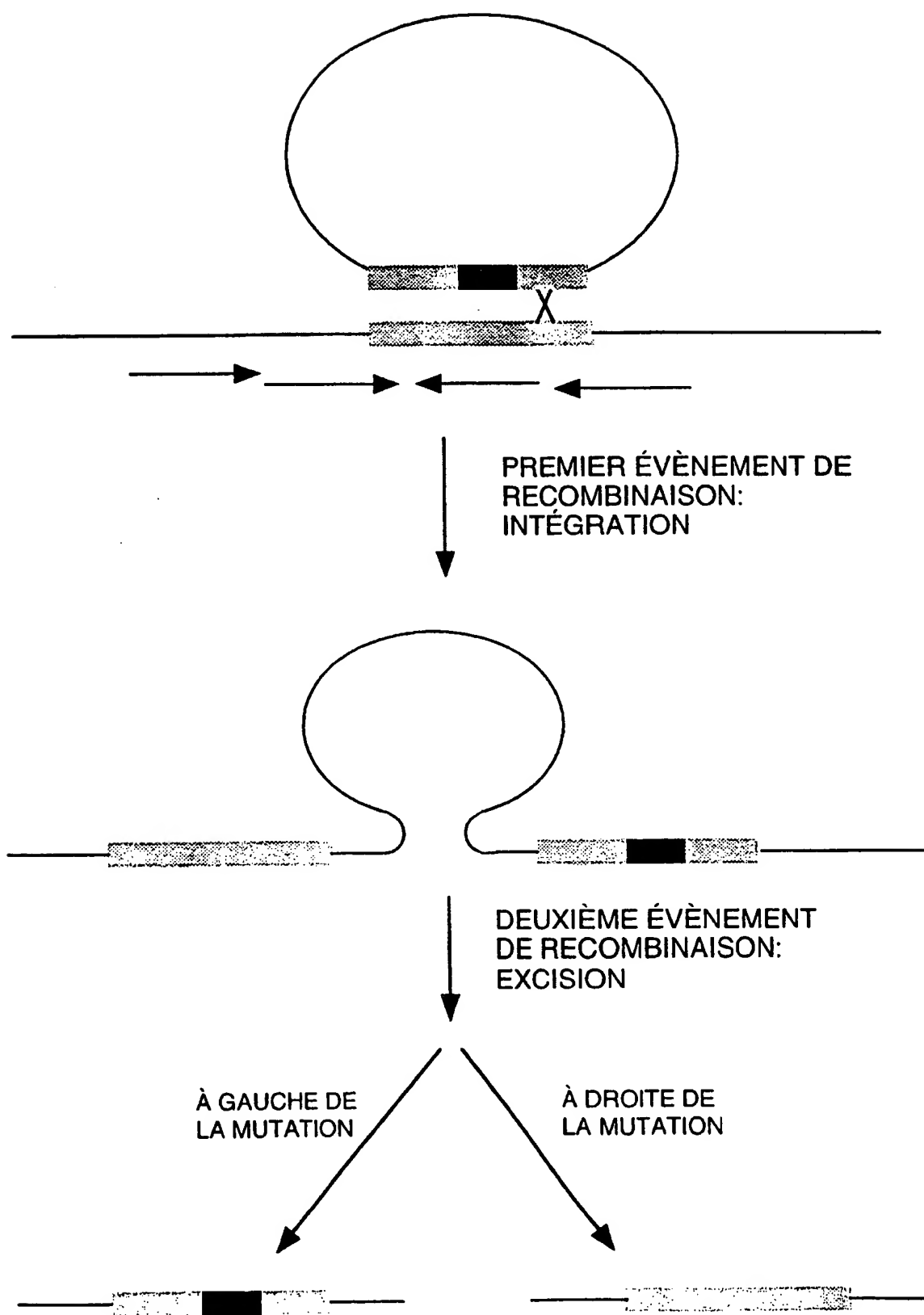


FIGURE 4

20/60

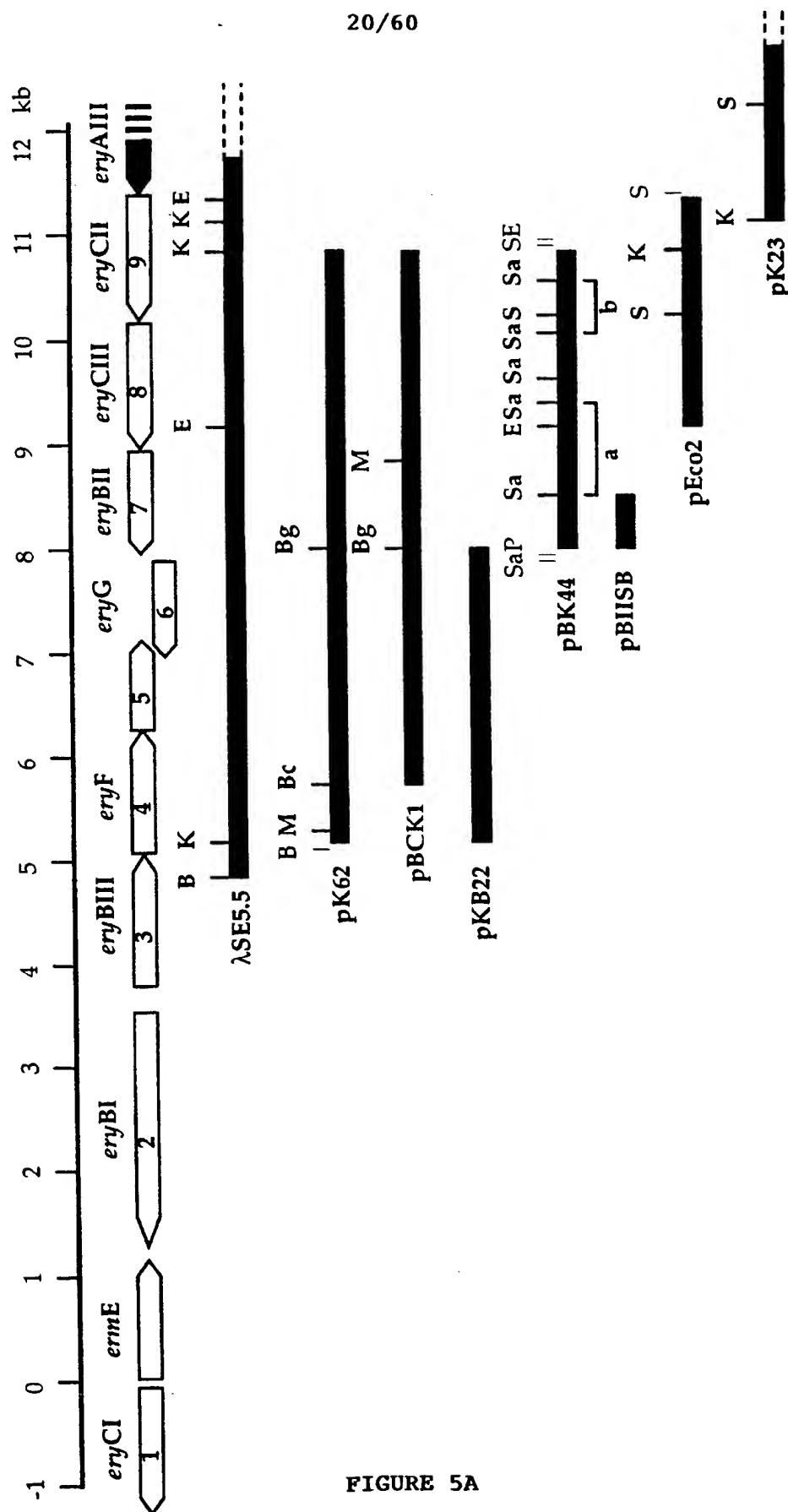


FIGURE 5A

21/60

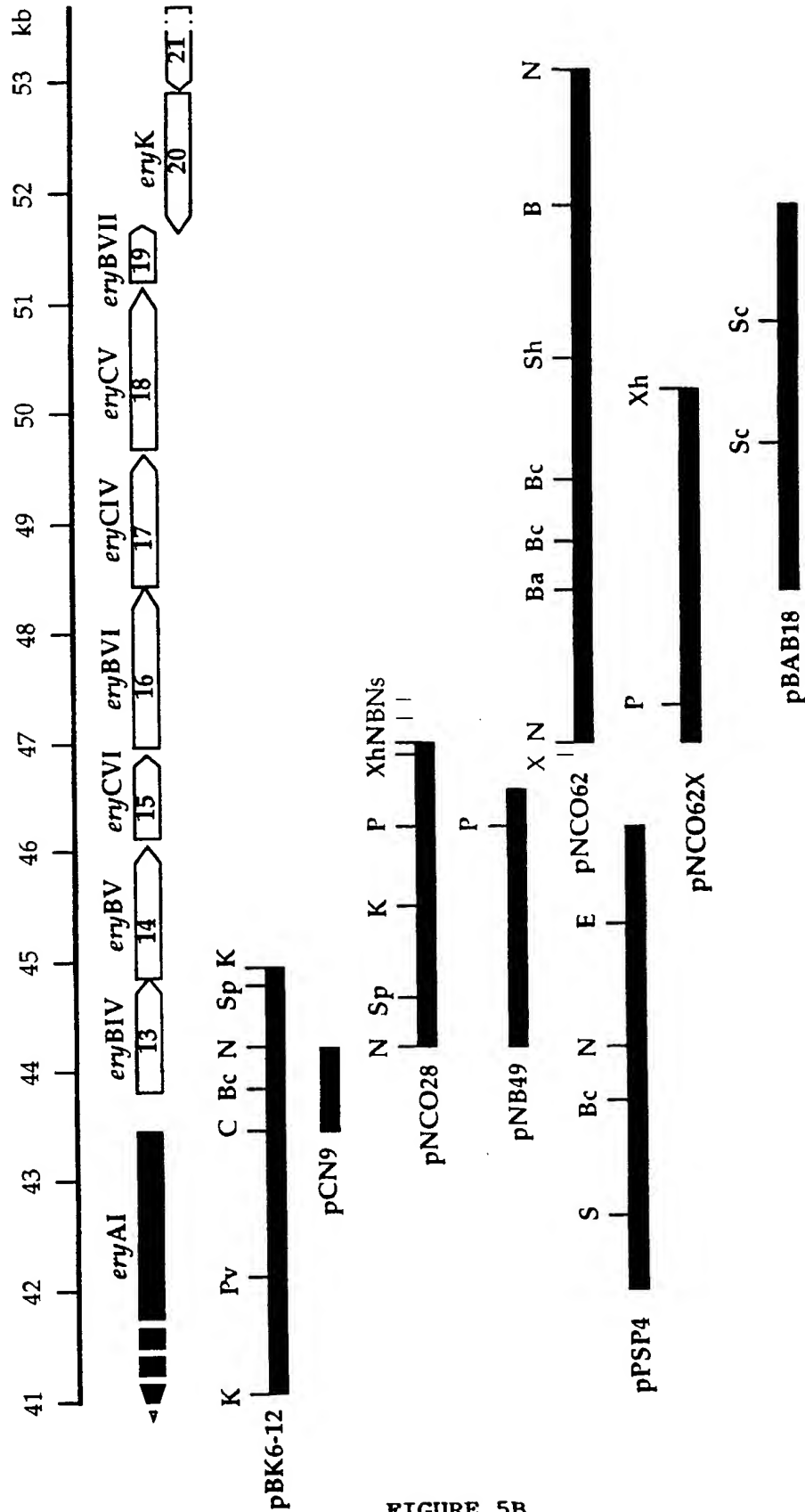


FIGURE 5B

22/60

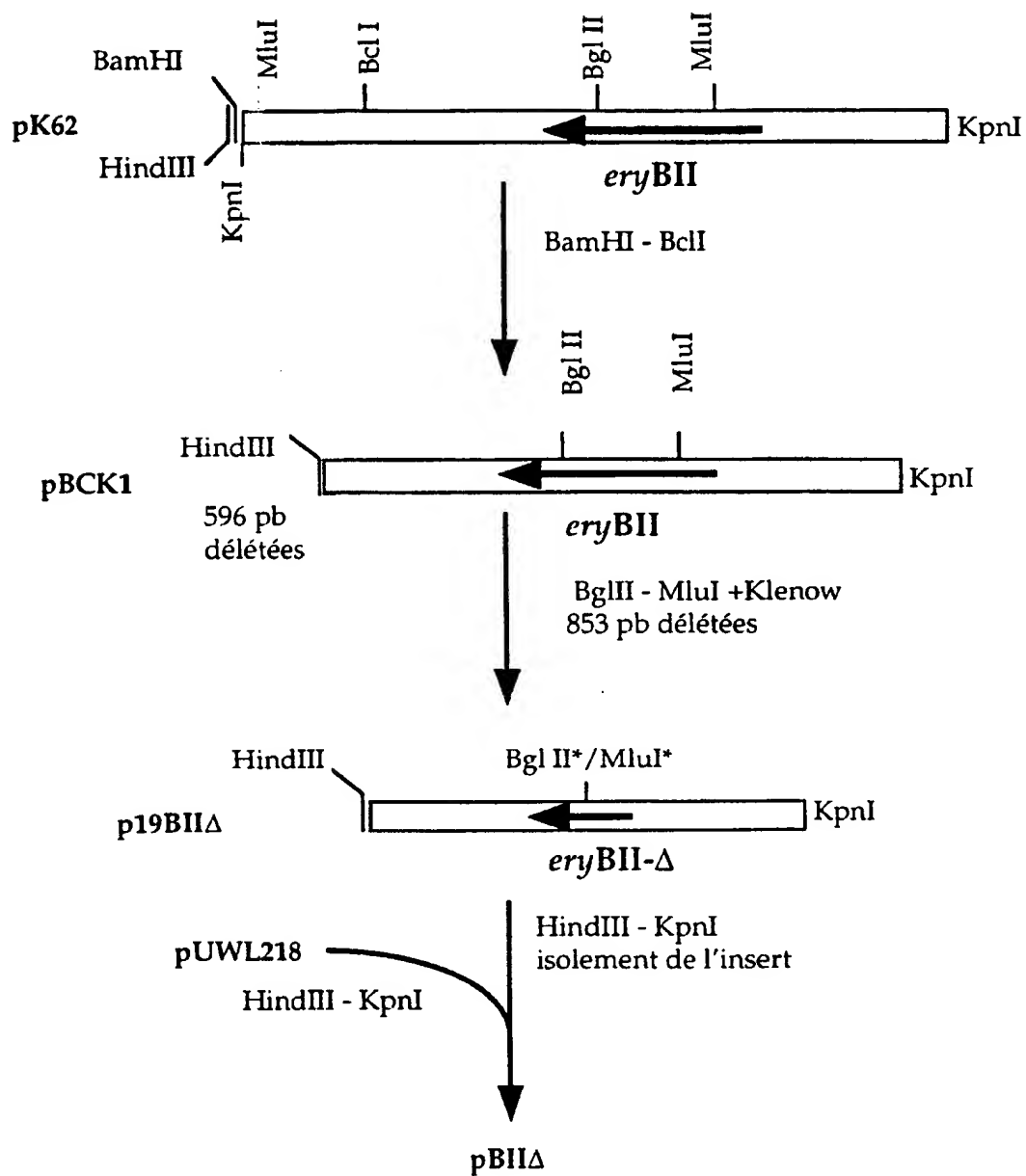


FIGURE 6A

23/60

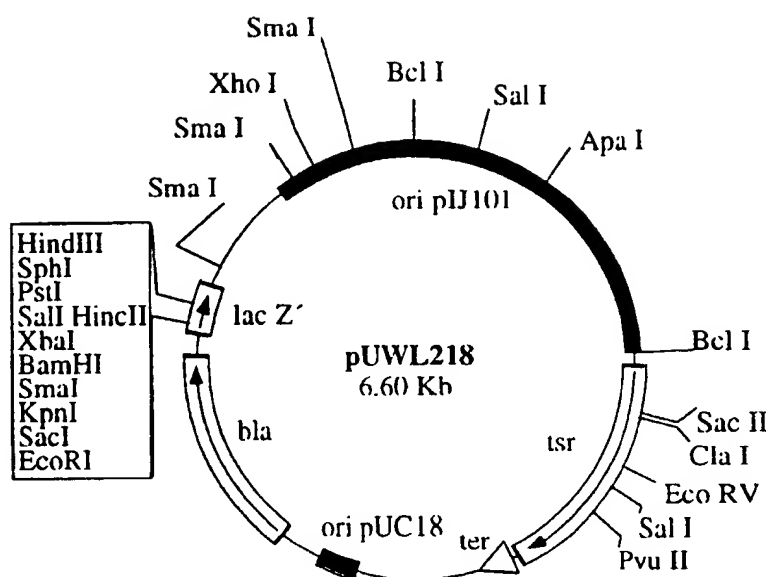


FIGURE 6B

24/60

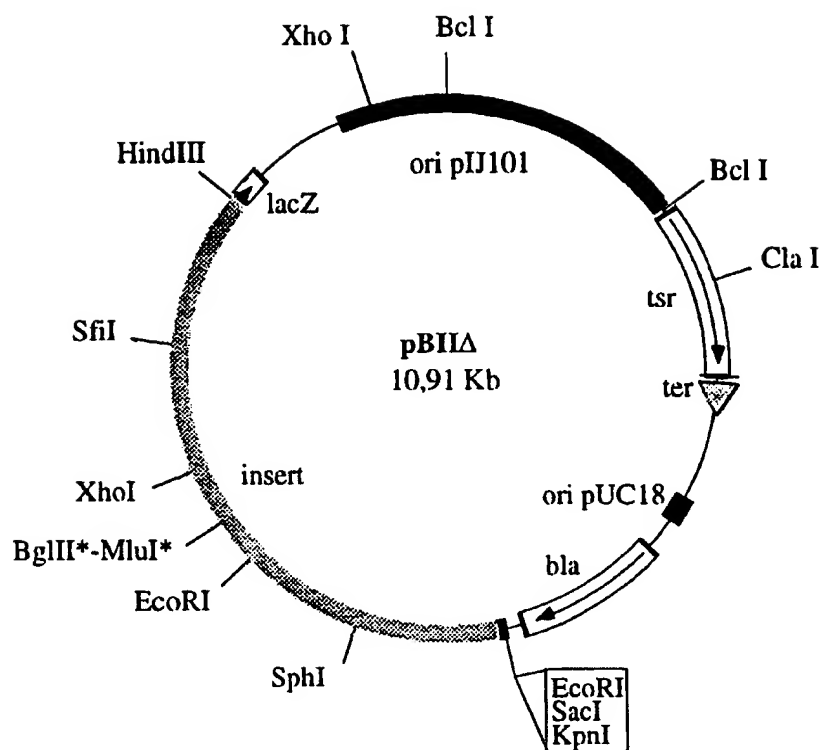


FIGURE 6C

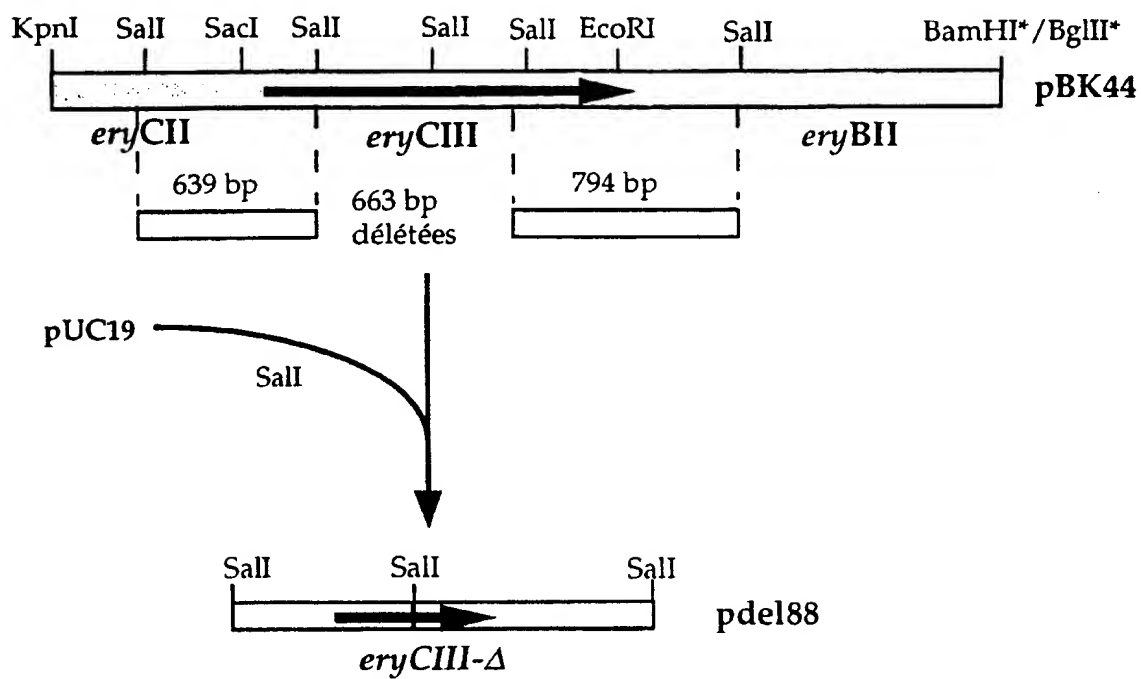


FIGURE 7A

26/60

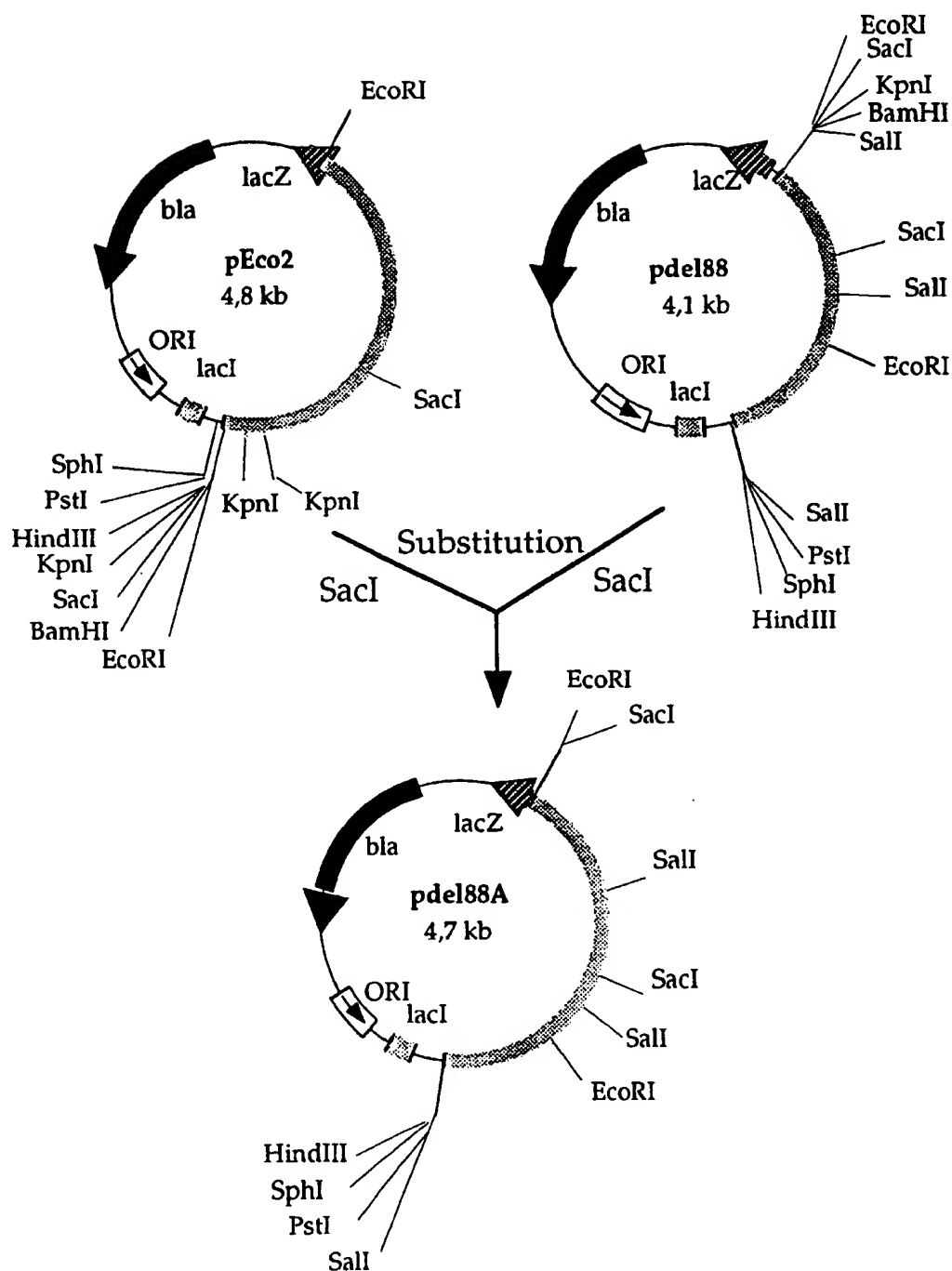


FIGURE 7B

27/60

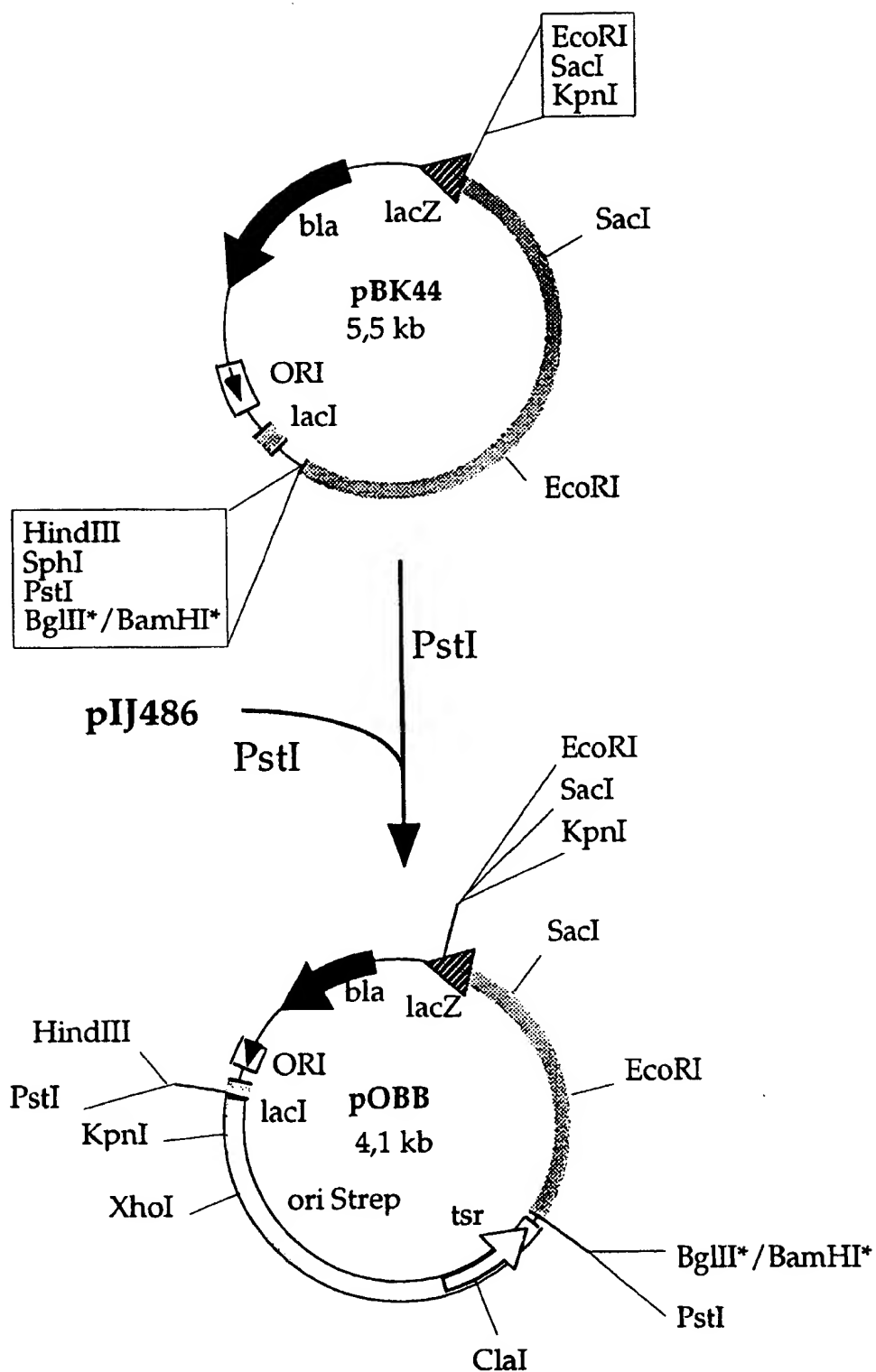


FIGURE 7C

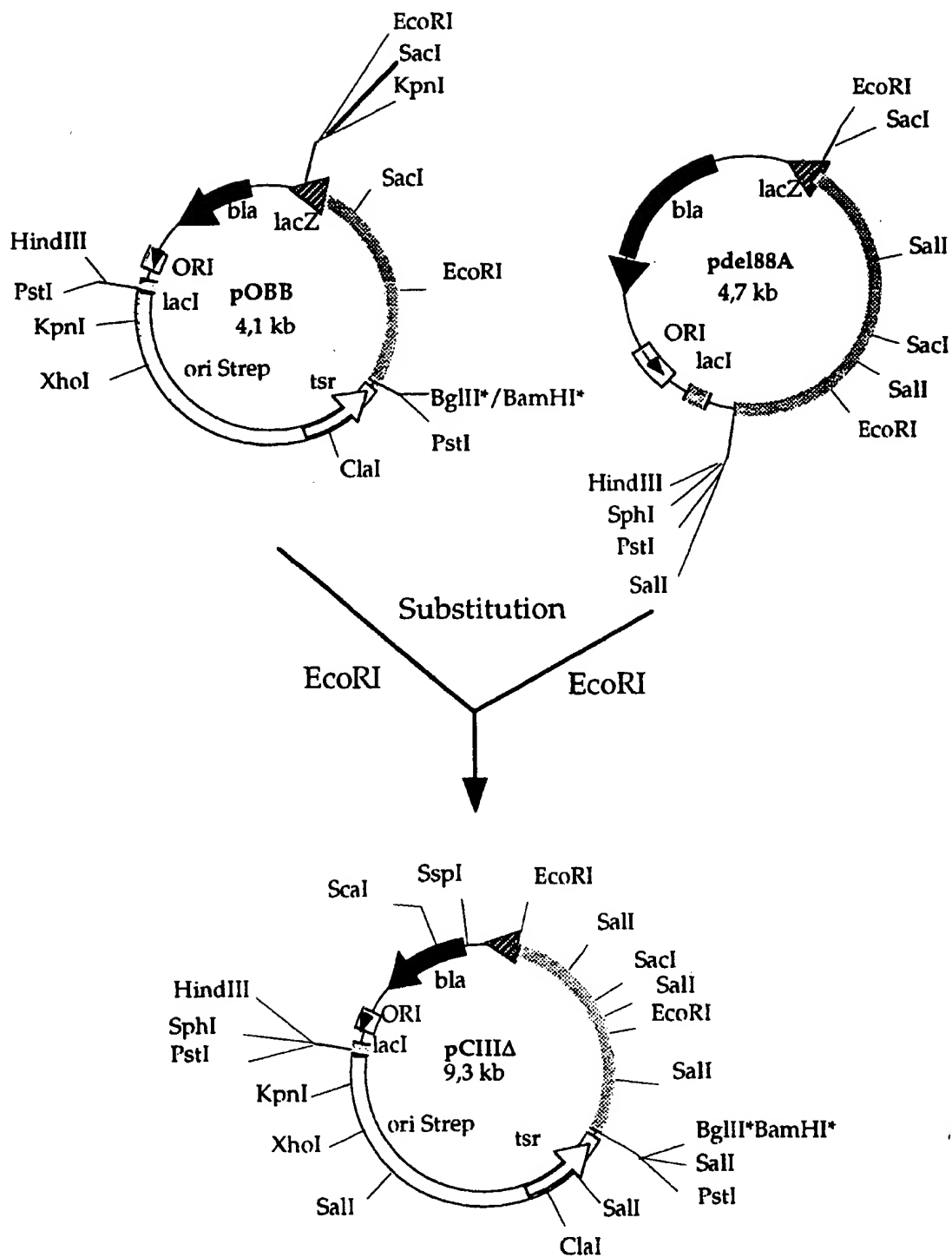


FIGURE 7D

29/60

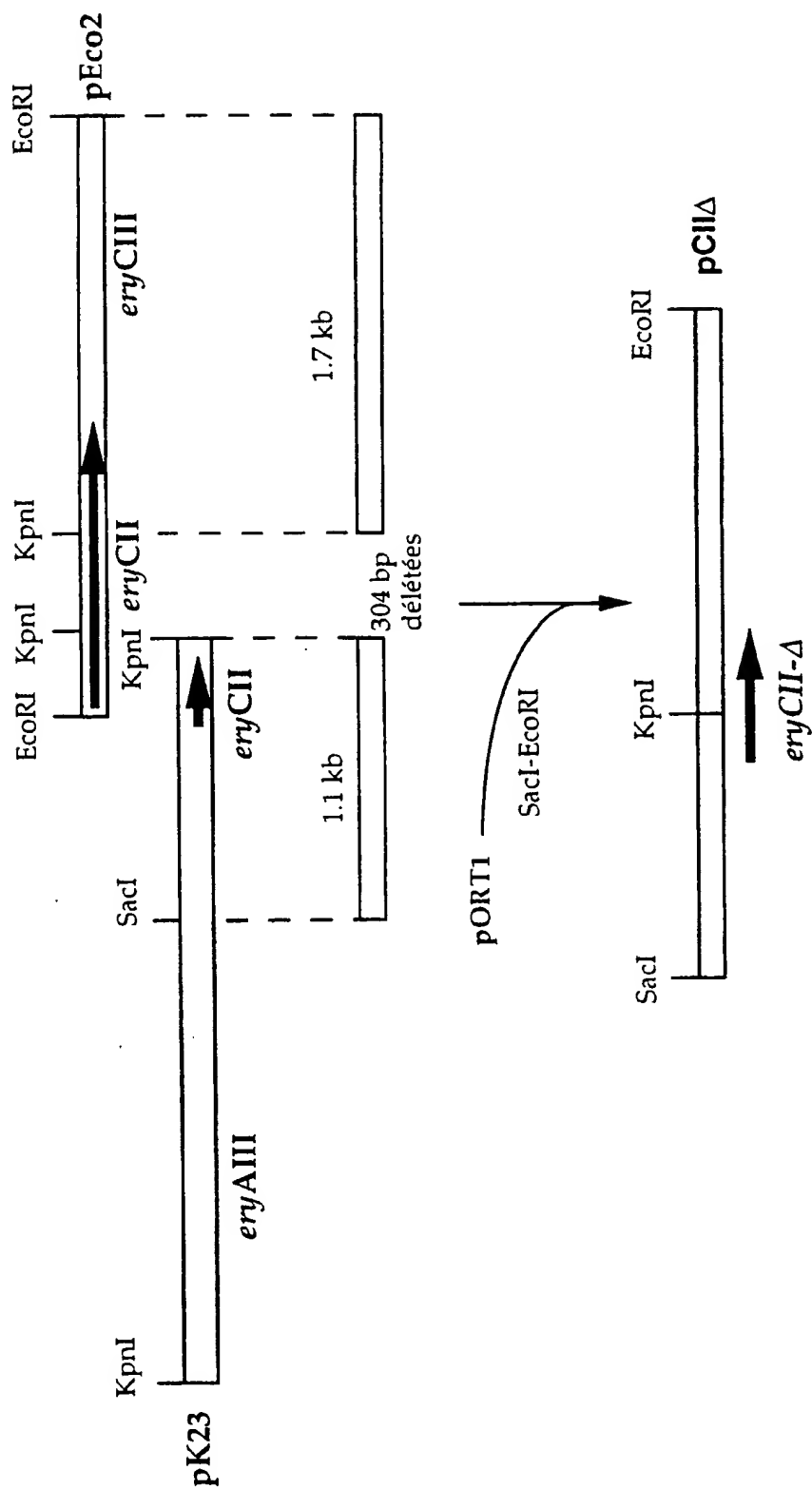


FIGURE 8A

30/60

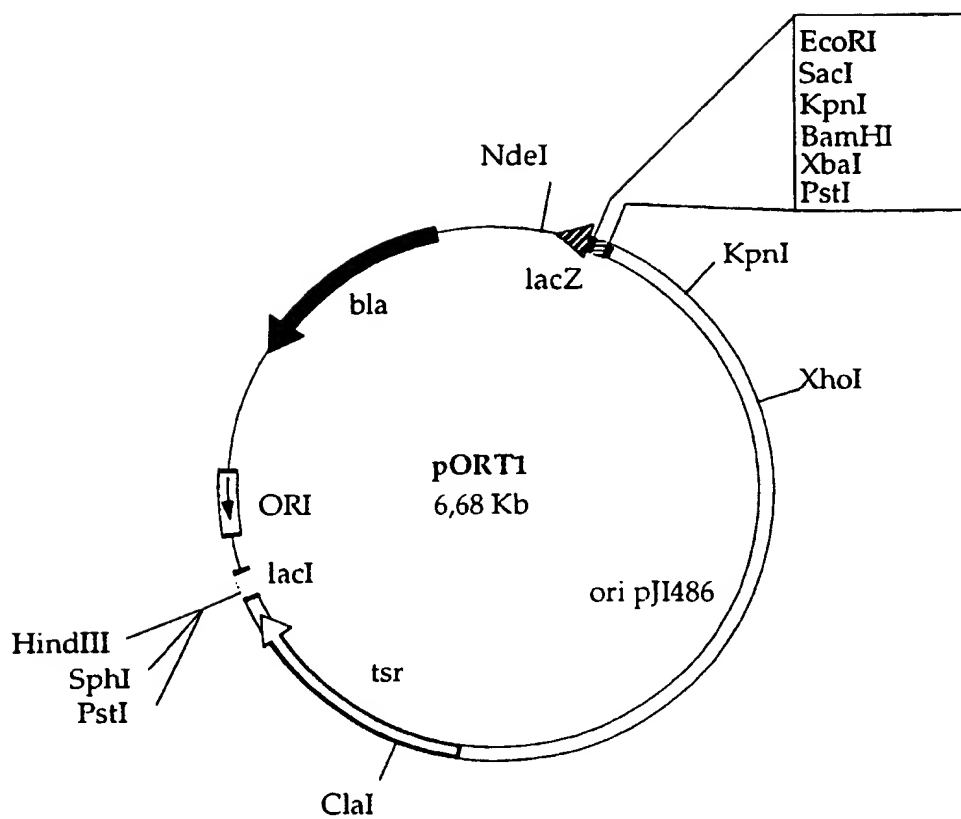


FIGURE 8B

31/60

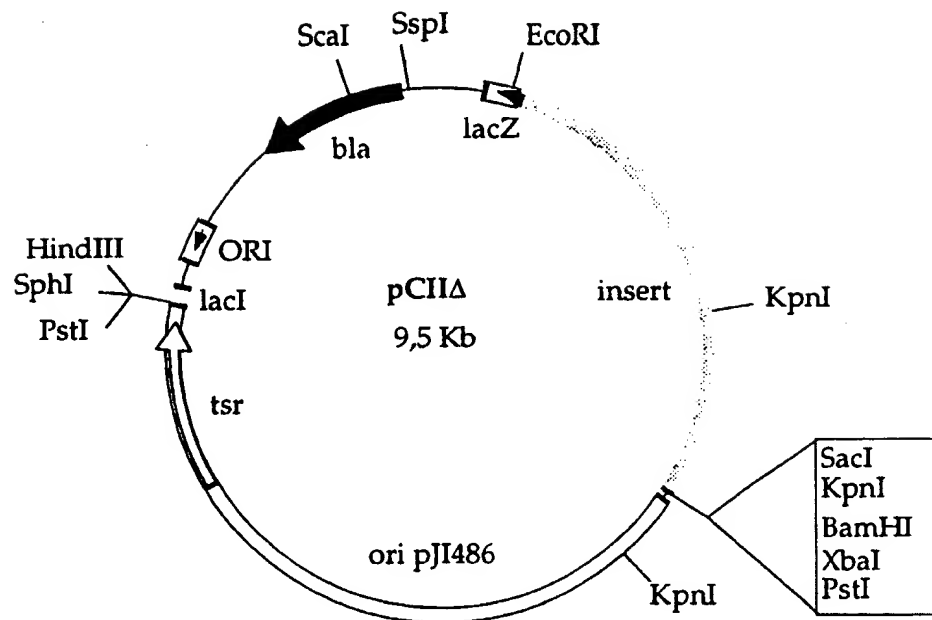


FIGURE 8C

32/60

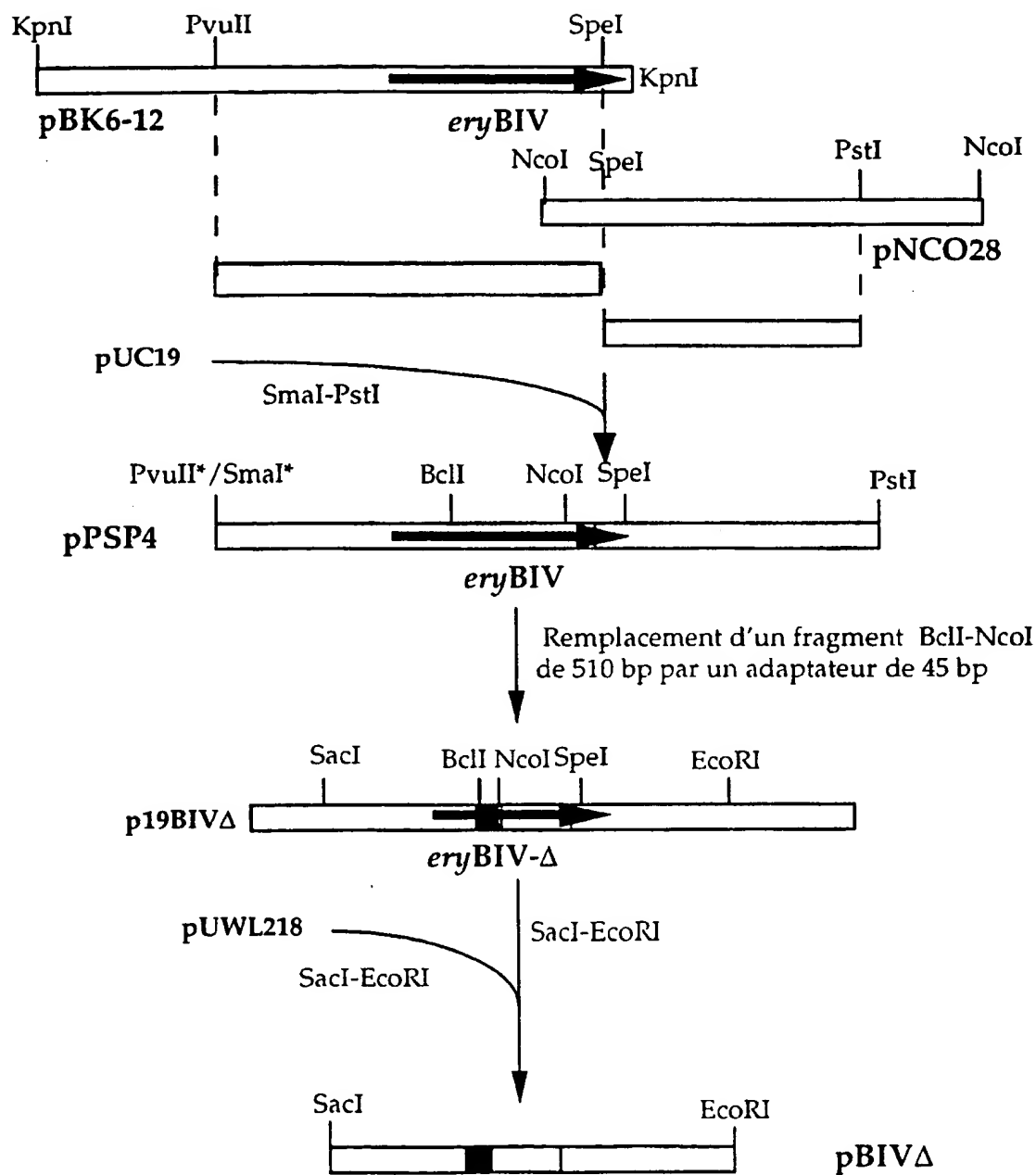


FIGURE 9A

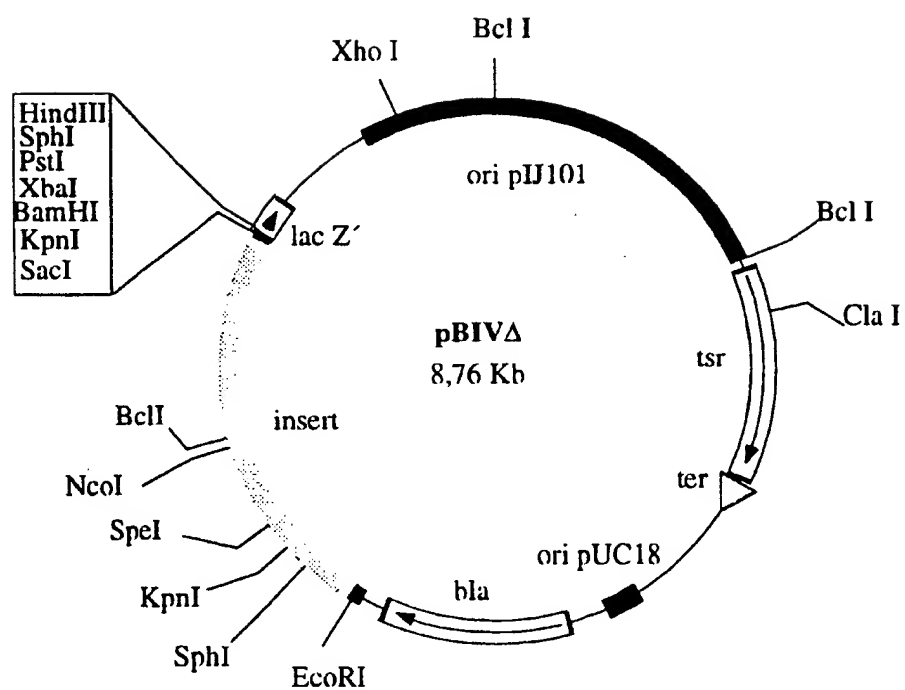


FIGURE 9B

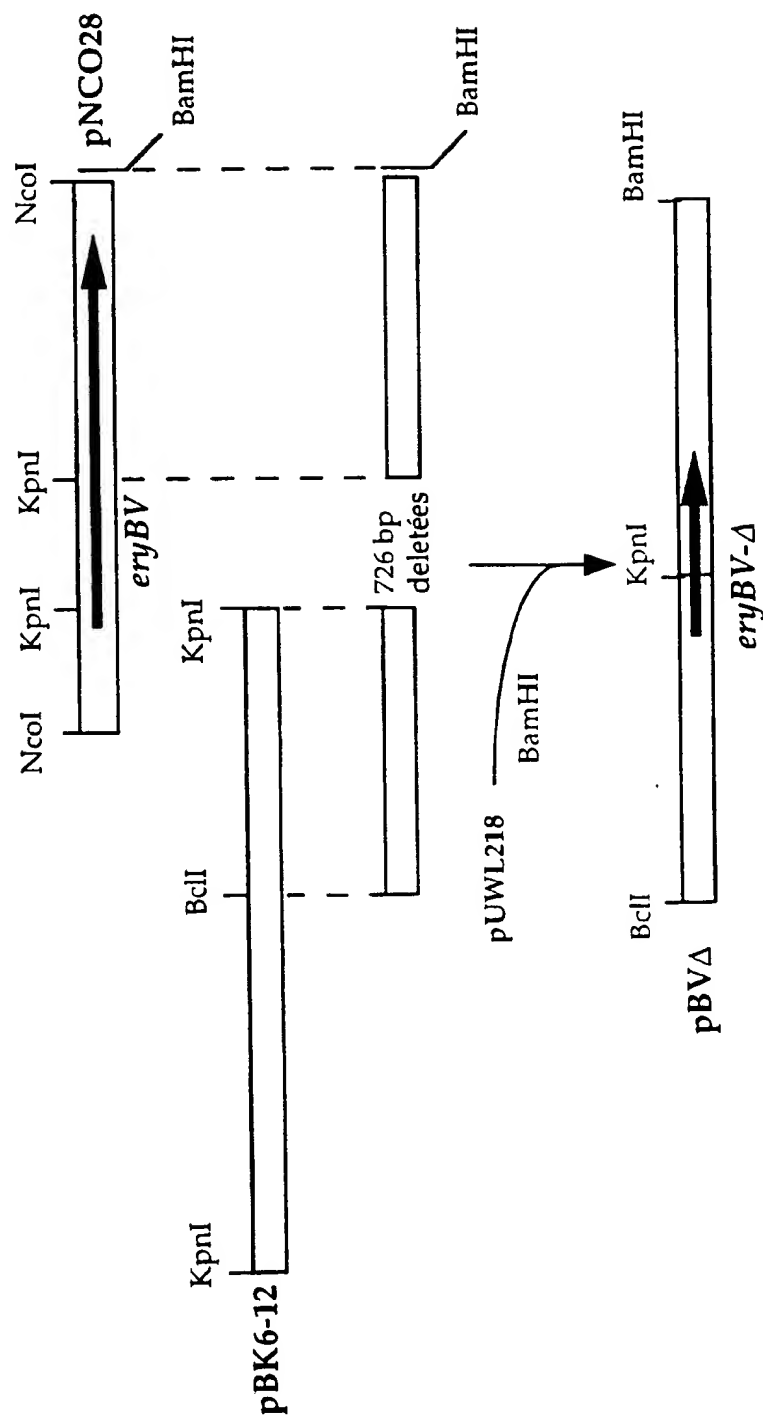


FIGURE 10A

35/60

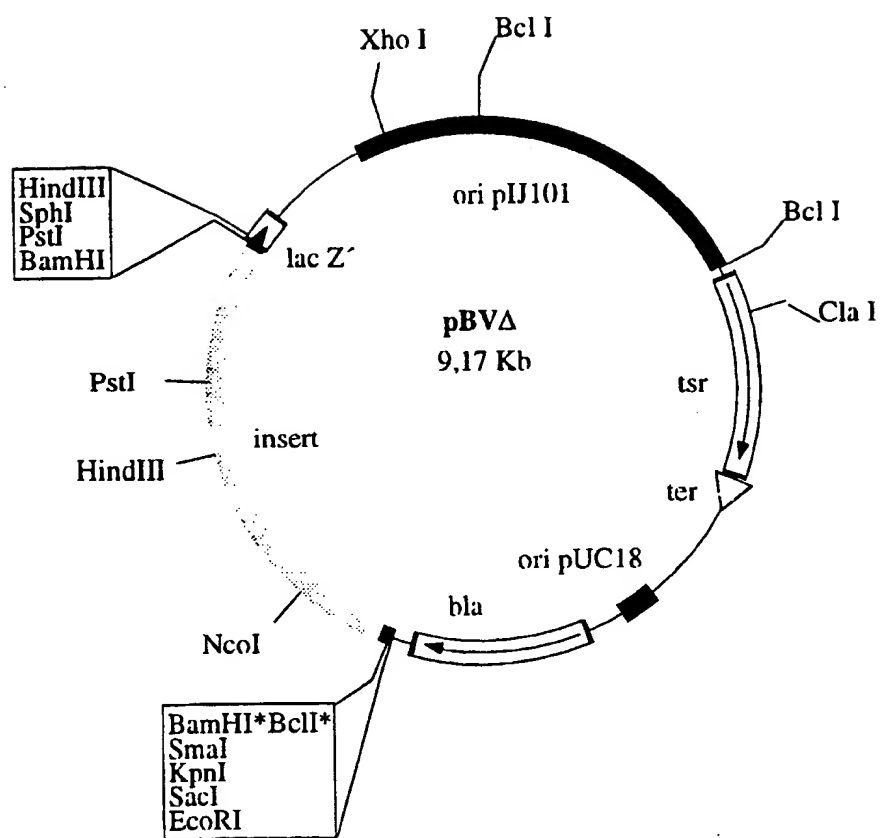


FIGURE 10B

36/60

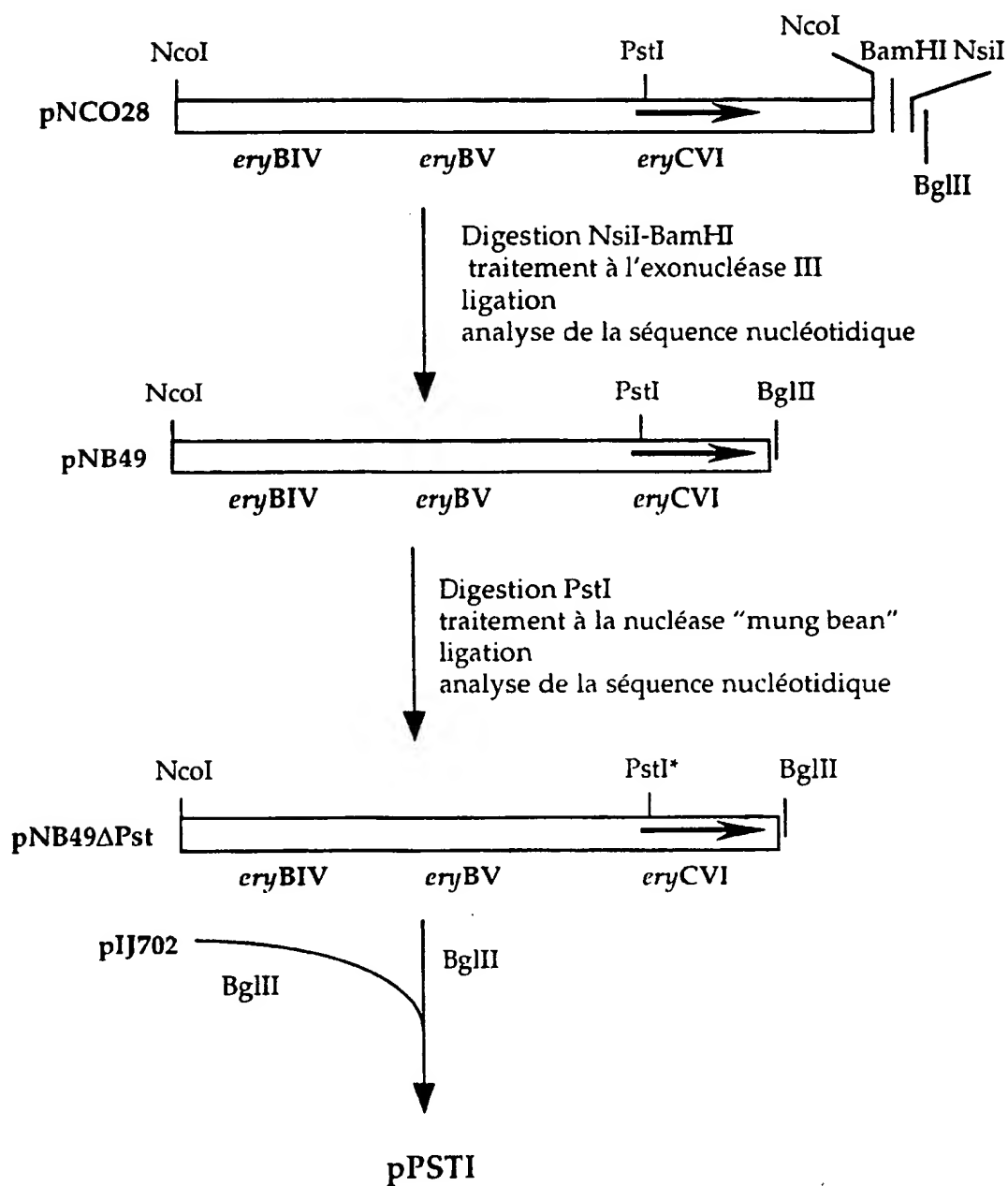


FIGURE 11A

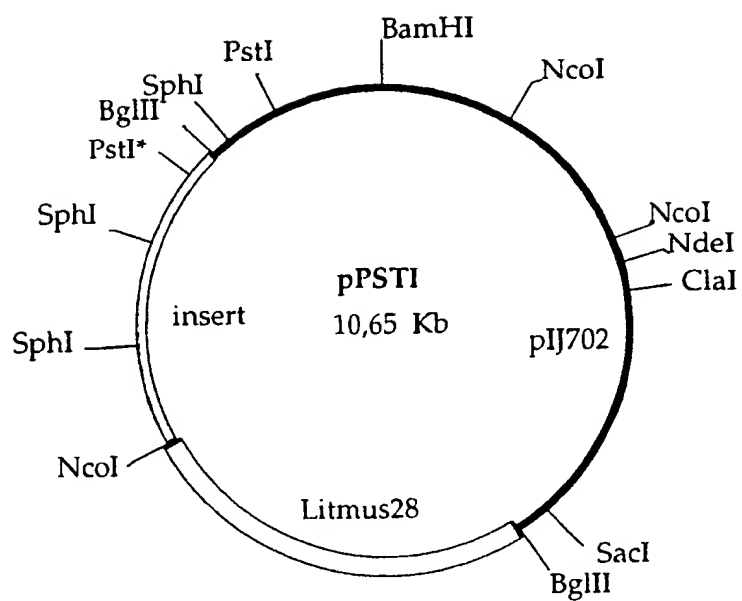


FIGURE 11B

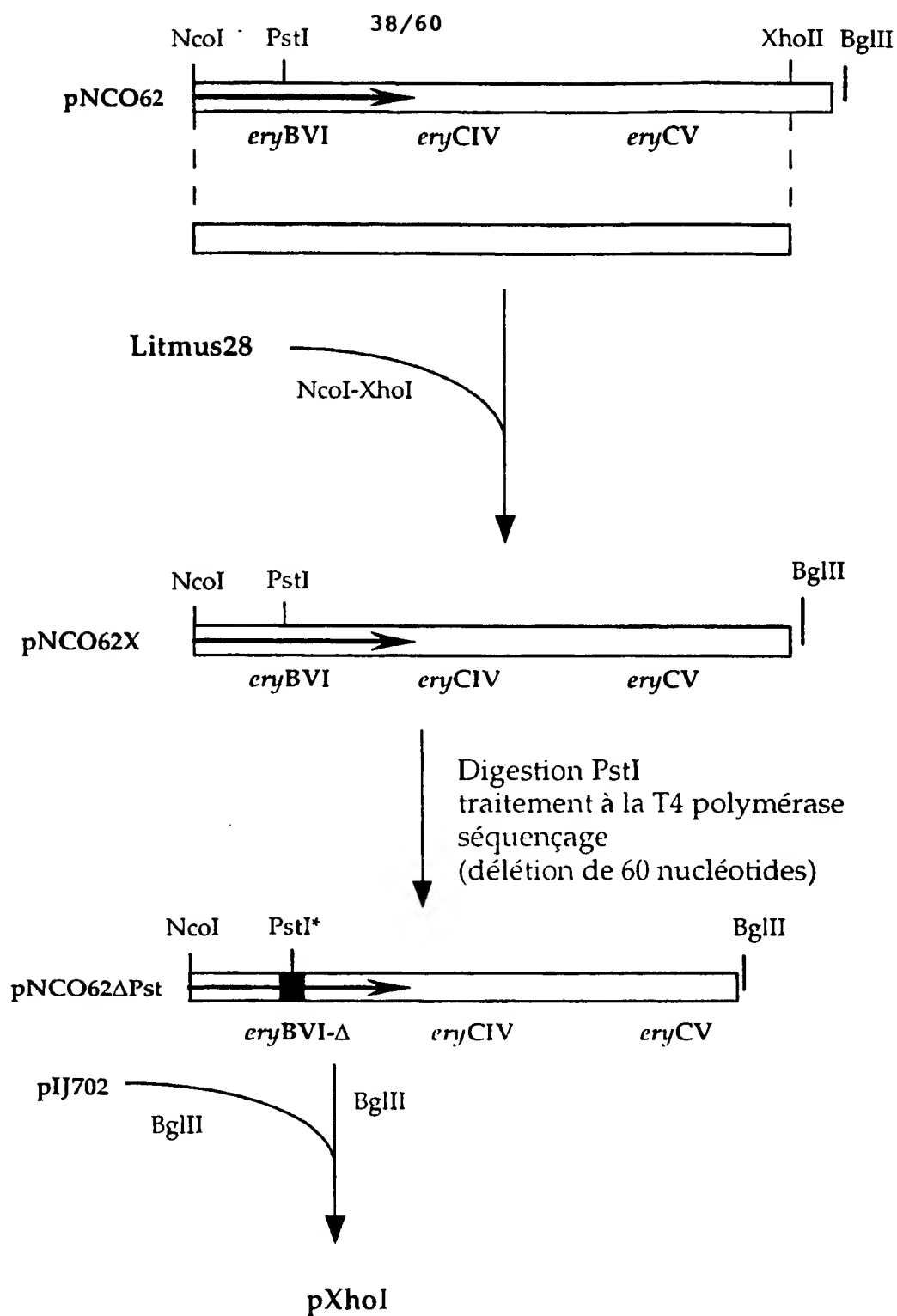


FIGURE 12A

39/60

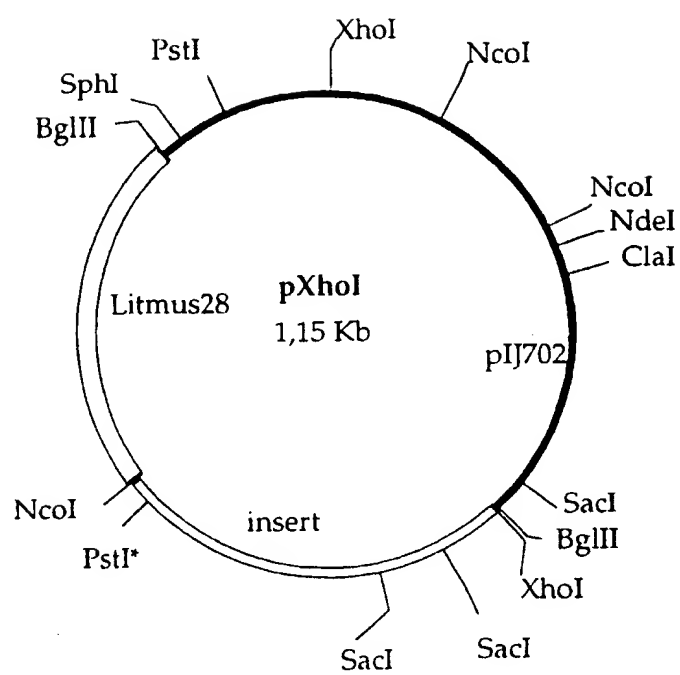


FIGURE 12B

40/60

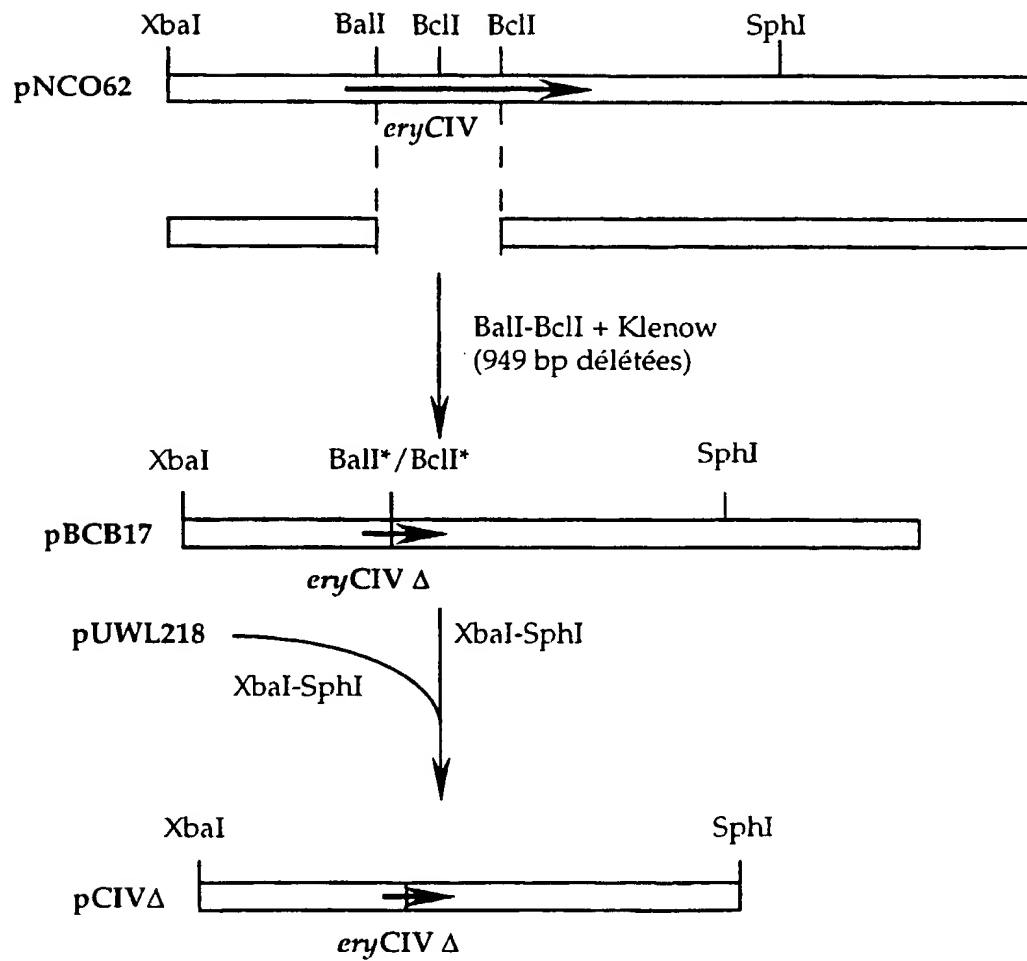


FIGURE 13A

41/60

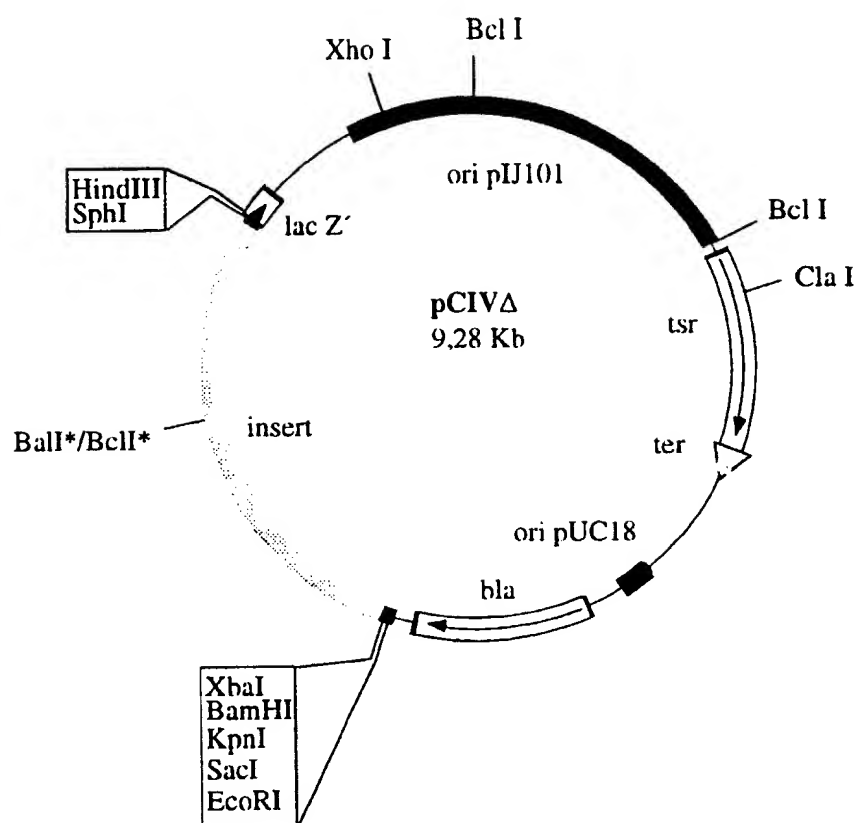


FIGURE 13B

42/60

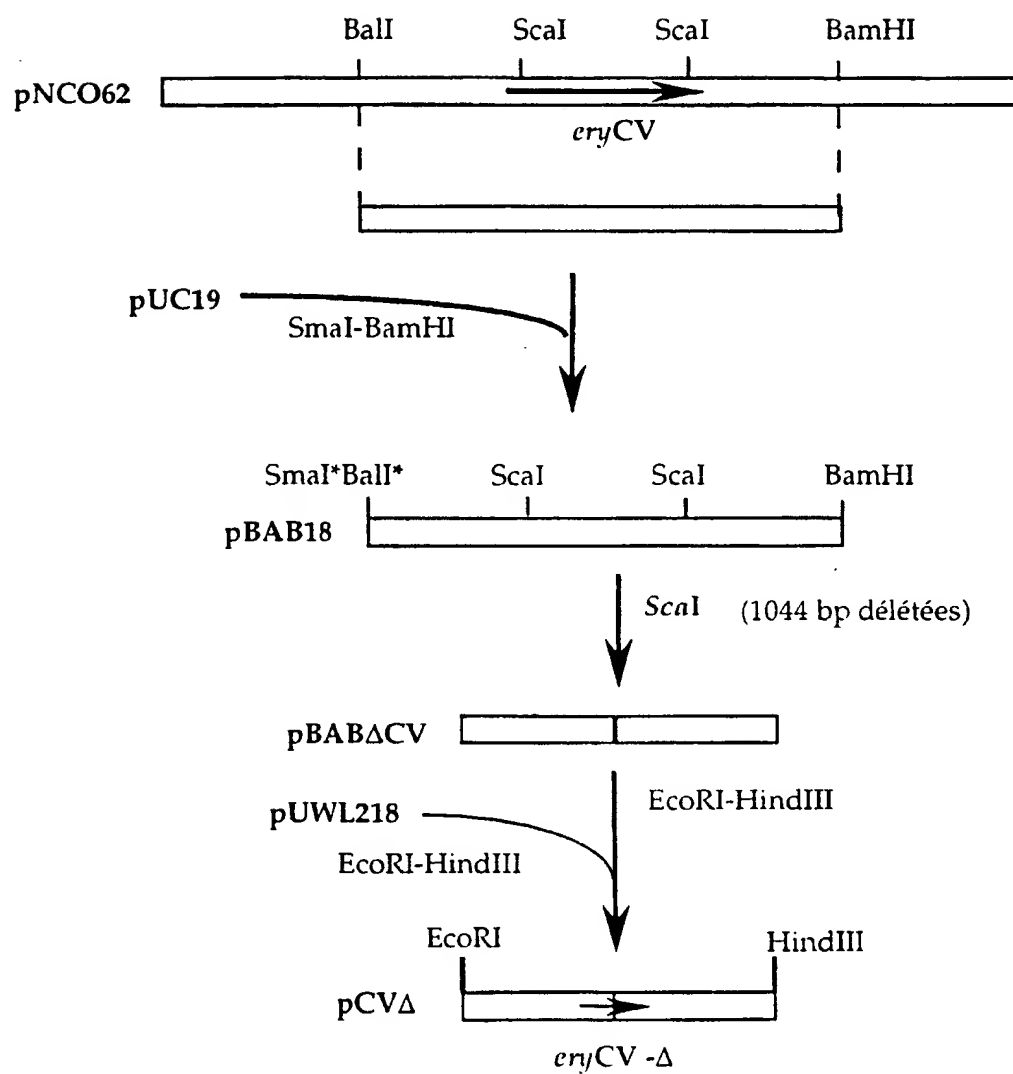


FIGURE 14A

43/60

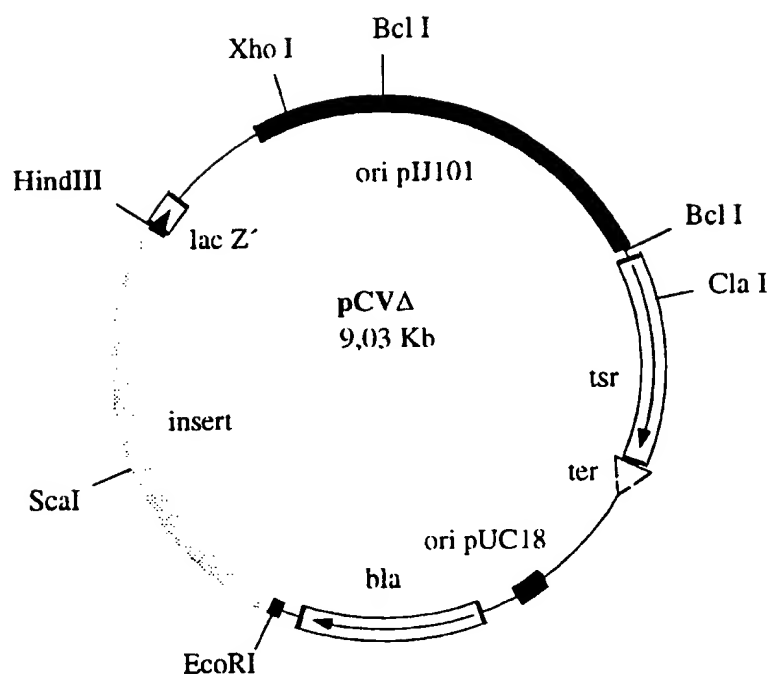


FIGURE 14B

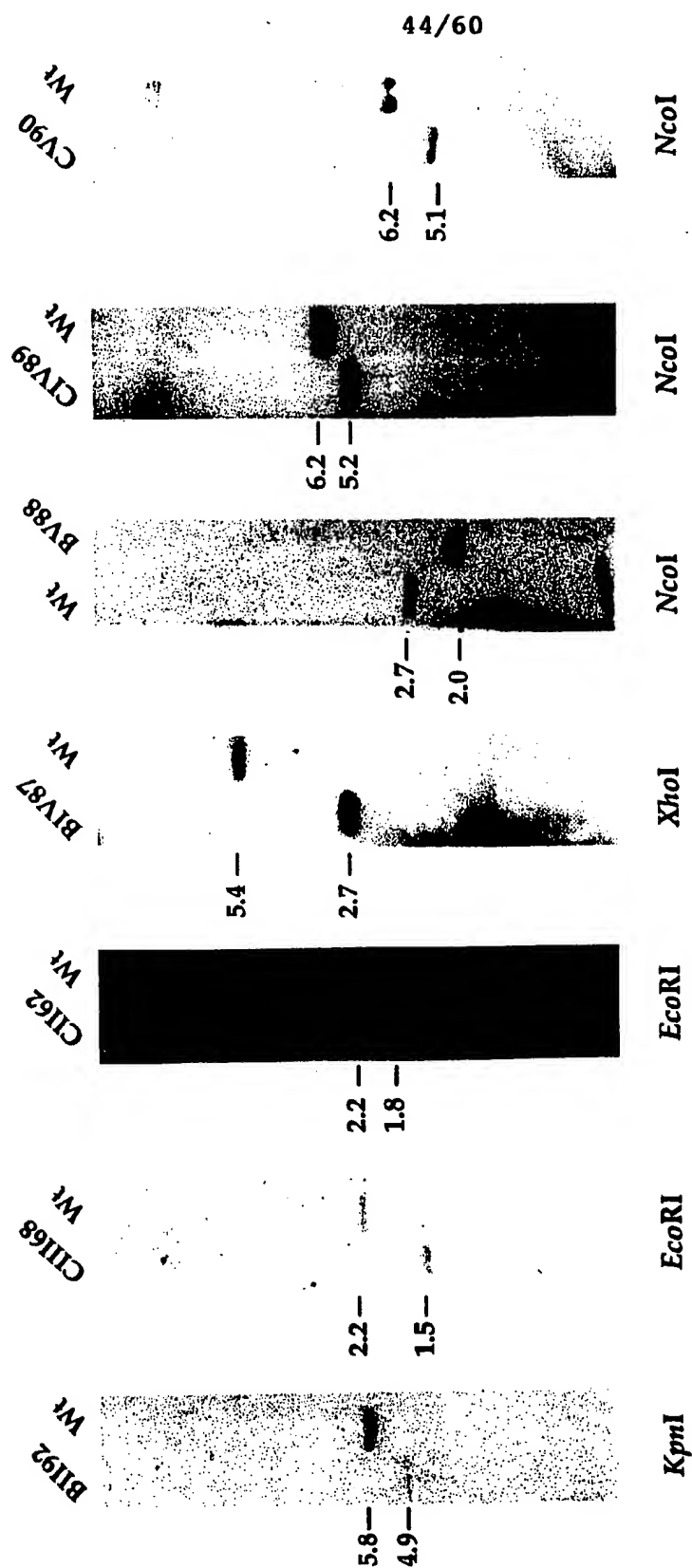


FIGURE 15

45/60

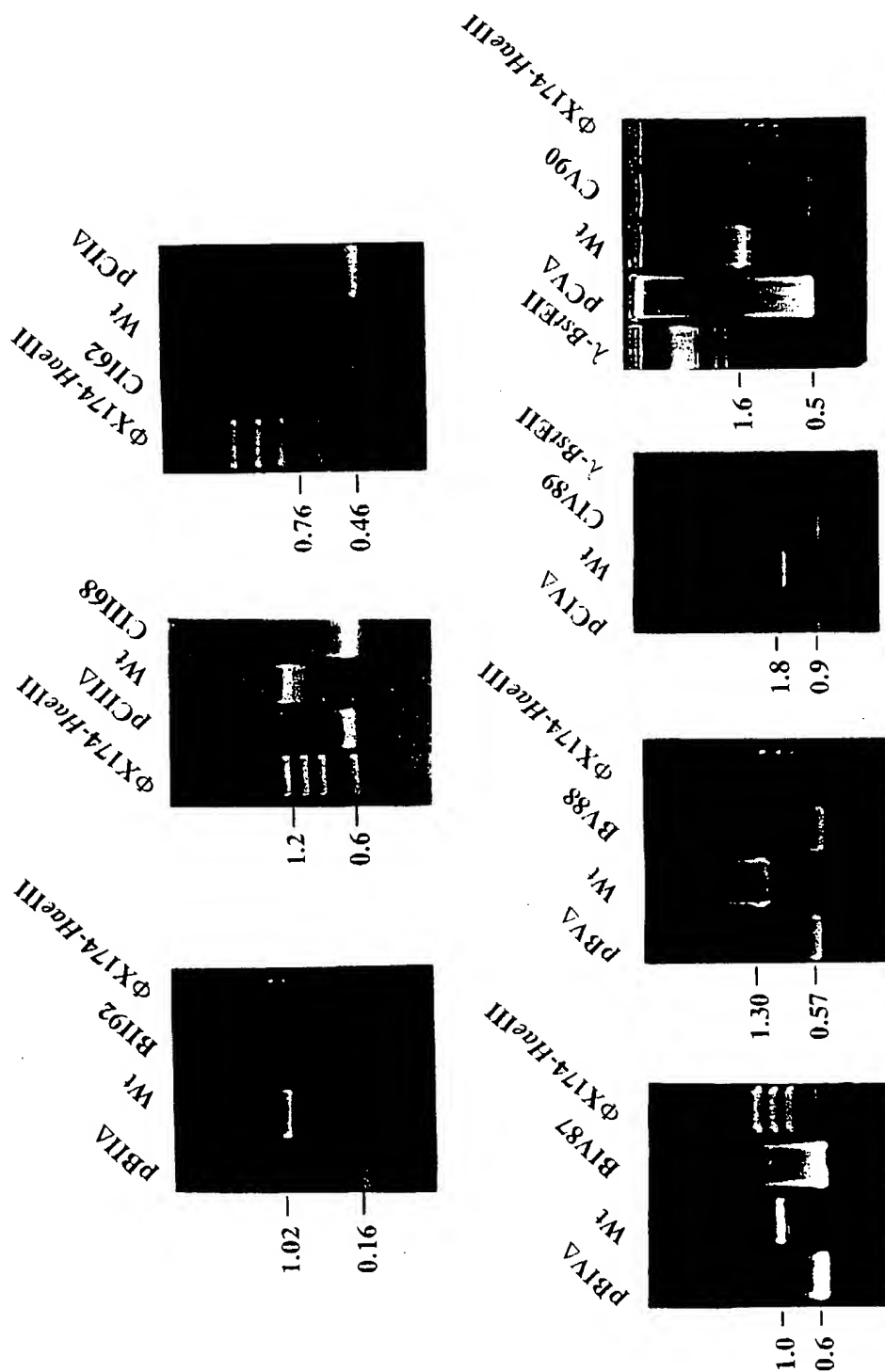


FIGURE 16

46/60

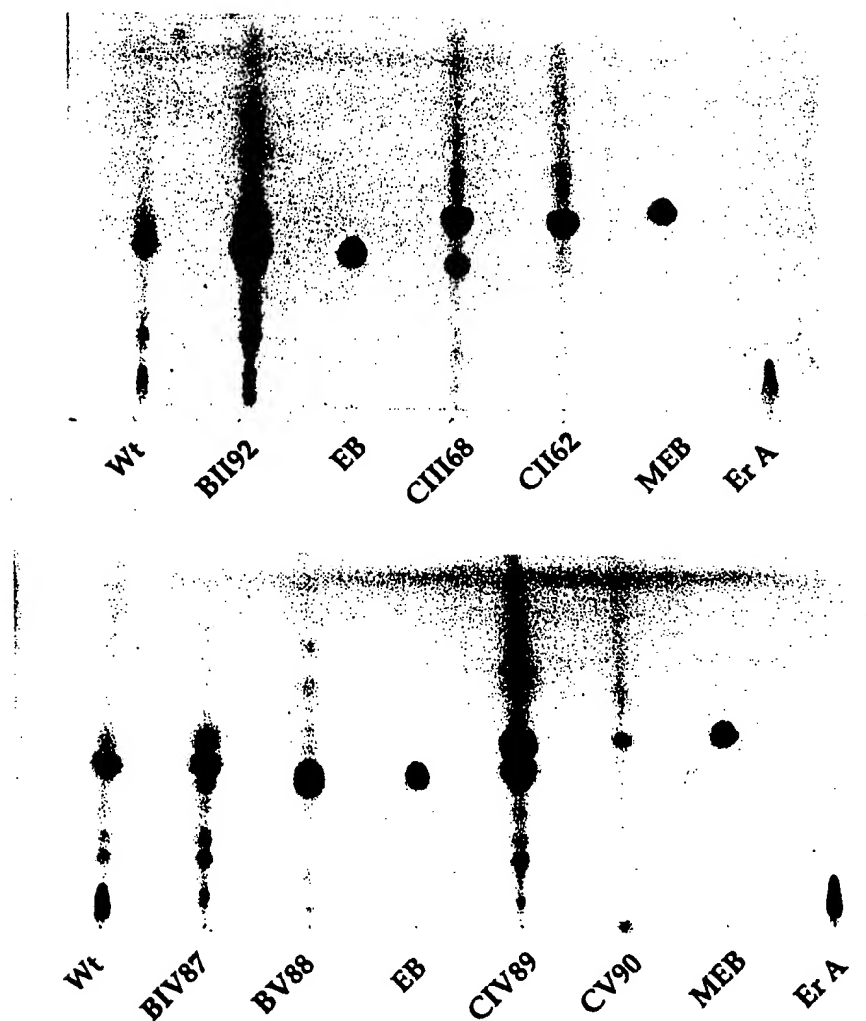


FIGURE 17

47/60



FIGURE 18

48/60

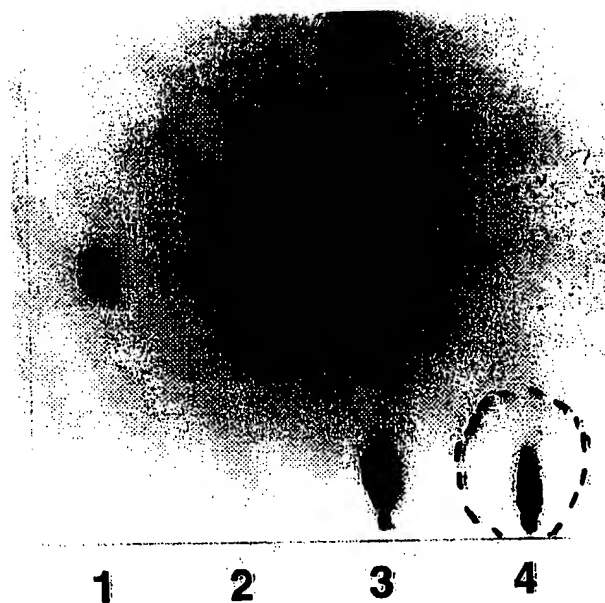


FIGURE 19

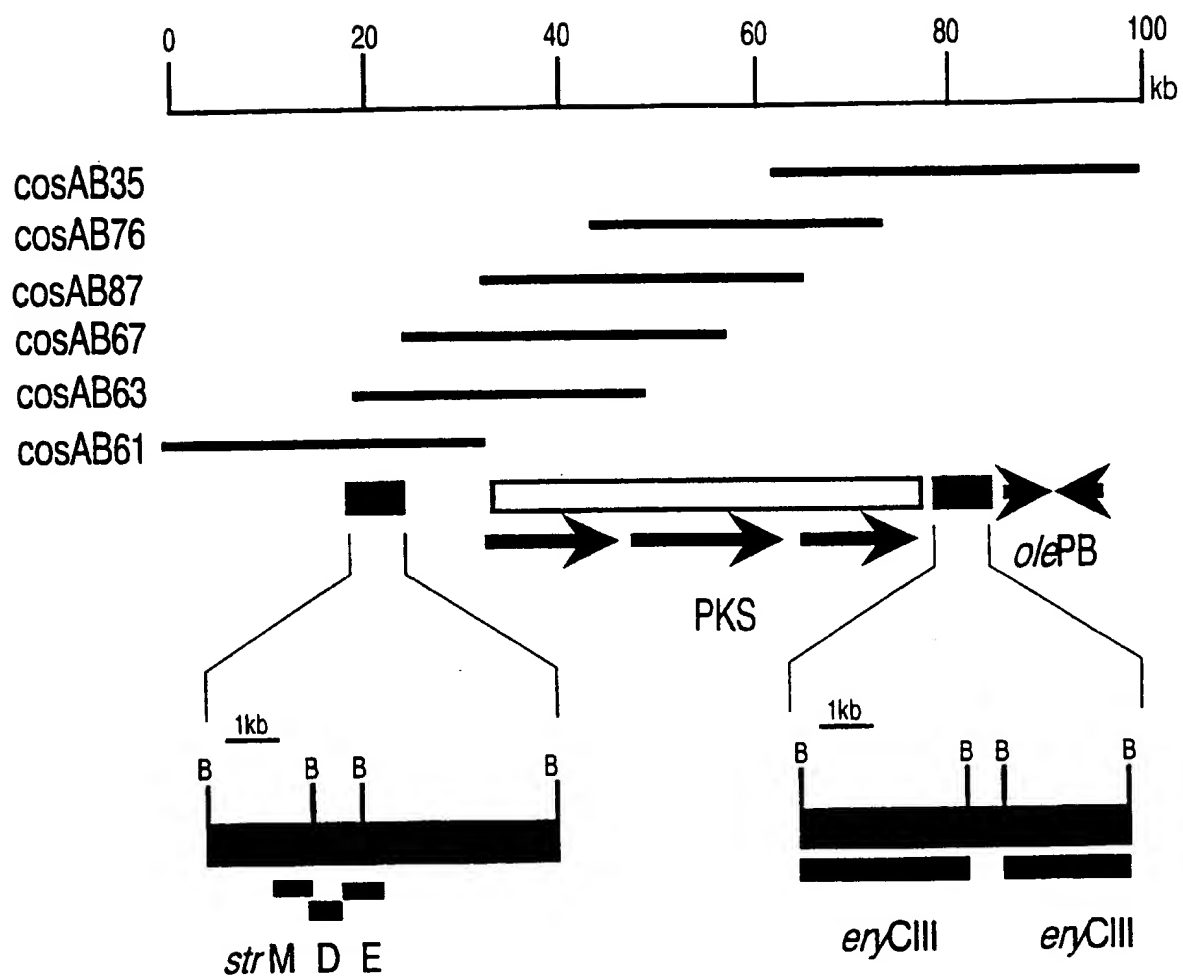


FIGURE 20

50/60

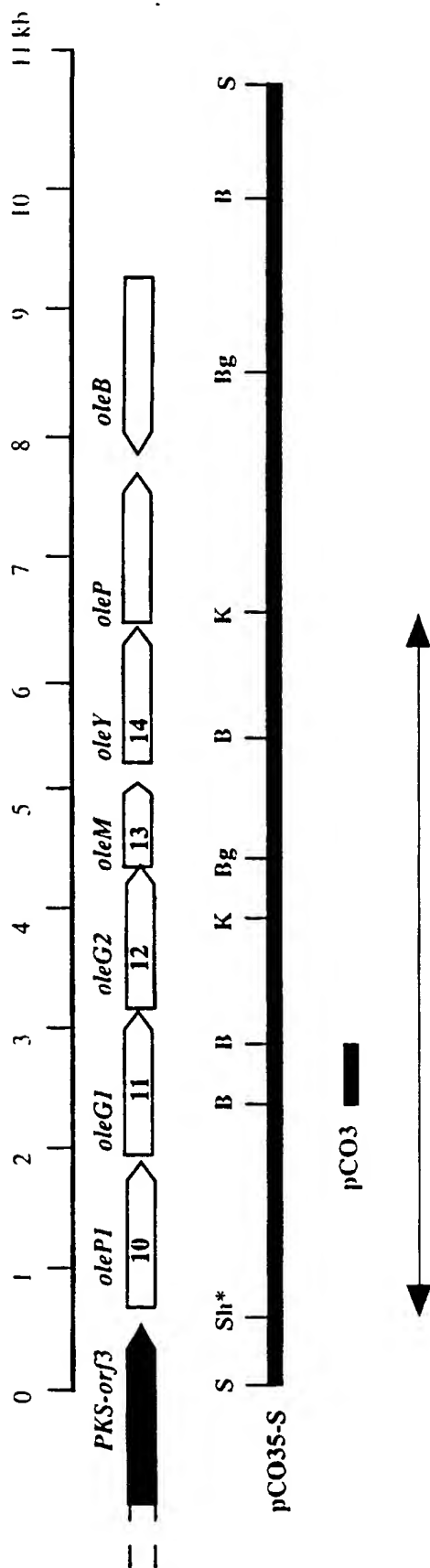


FIGURE 21

```

1  GCATGCCCCGCTTTCCCTCCCCCTCTCCGAACGCATCGACGACCCGATCCCCCTCAGGGAC  60
61  CGGTGAAGGAGCGTGTGCACTCATGAGGACATGCAAGCGGTACAGCCCGAACCAGCCA  120
121 GTGTCGAACACGGGGGACCGCAGCTCGAACAGAGCGAACGGCGCACGGAAGCCGCCAG  180

                                     PstI
181 GAGATGGAGGACAGCGAACTGGGGCGCCGCTGCAGATGCTCCGGCGCATGCAGTGGGTC  240
    M E D S E L G R R L Q M L R G M Q W V
    OleP1
241 TTCGGCGCCCAACGGCGGATCCGTACGCCCGGCTGCTGTGTGGCATGGAGGATGACCCGTCA  300
    F G A N G D P Y A R L L C G M E D D P S
301 CCTTTCTACGACGCGATACGGACCCCTGGCGGAGCTGCACCGGAGCAGGACCGGAGCCTGG  360
    P F Y D A I R T L G E L H R S R T G A W
361 GTCACCGCGACCCCGGCTCGGGGGCCGCTCCTCGCCGACCGGAAGCTCGGTGCCCG  420
    V T A D P G L G G R I L A D R K A R C P
421 GAAGGCTCGTGGCCGGTGGGGCGAAGACCGACGGGCTGGAGCAGTACGTGCTGCCCGGG  480
    E G S W P V R A K T D G L E Q Y V L P G
481 CACCAGGCGTTCTGCGGCTGGAGCGCGAGGAGCGCCGAGCGACTGCGGGAGGTGCGCGCG  540
    H Q A F L R L E R E E A E R L R E V A A
541 CCGGTGCTGGGGCGCGCGGTGACCGGTGGCGCCCGCTGATCGACGAGGTCTGCGCG  600
    P V L G A A A V D A W R P L I D E V C A
601 GGGCTCGGAAGGGGCTGCCGGACACGTTTCGACCTGGTCGAGGAGTACGCGGGGCTGGTG  660
    G L A K G L P D T F D L V E E Y A G L V
661 CCGGTCGAGGTGCTGGCGCGGATCTGGGGCGTCCCGGAGGAGGACCGCGCCCGGTTCCGG  720
    P V E V L A R I W G V P E E D R A R F G

```

FIGURE 22

721 CGTGA CTGCGCGGCTCGCTCCCGCGCTGGACAGCCTCCTGTGTCCCCAGCAGTTGGCG 780
R D C R A L A P A L D S L L C P Q L A

781 CTGAGCAAGGACATGGCGTCCGCCCTGGAGACCTGCGTCTCCTCTTCGACGGCCCTCGAC 840
L S K D M A S A L E D L R L L F D G L D

841 GCGACGCGCGCTCGCGGCGCCCGCGGACGGTGACGGAACGGCCGTGGCCATGCTCACC 900
A T P R L A G P A D G D G T A V A M L T

901 GTTCTGCTCTGCA CGGAGCCGGTGACCA CGGCGATCGGGAACACCGTGCTCGGGCTCCTT 960
V L L C T E P V T T A I G N T V L G L L

961 CCCGGGAGTGGCCCGTGCCTGCACCGGCGGGTGCTGCCGGCAGGTTGCCGGGCAG 1020
P G Q W P V P C T G R V A A G Q V A G Q

1021 GCGCTGCACCGGCGGTGTGTA CCGTATCGCGACCGCGSTTCGCCCGGAGGACCTGGAG 1080
A L H R A V S Y R I A T R F A R E D L E

1081 TTGGCGGGCTGCGAGGTCAAGTCCGGTGACGAGGTGGTGGTCTGCGCGGAGCGATCGGC 1140
L A G C E V K S G D E V V L A G A I G

1141 CGGAACGGACCGTCCGCGAGCCCGCGCTGCCCGGCGCCAGCGGCGCCCGCCGCC 1200
R N G P S A A A P P A P P G P A A P P A

1201 CCGTCGGTCTTCGGTGCCGCGCCTTCGAGAACGGCTGGCCGAACCCCTCGTCCGGGCT 1260
P S V F G A A A F E N A L A E P L V R A

1261 GTGACGGGAGCGGCCCTCCAGGCCCTCGCGGAGGGGCCCCCGGCTGACGGCGGGGGA 1320
V T G A A L Q A L A E G P P R L T A A G

1321 CCCGTGCTACGACGGCGGGCTTCCCTGTGCTGCGGGGCTGCACCGGGCTCCGGTGGCC 1380
P V V R R R R S P V V G G L H R A P V A

FIGURE 22

FIGURE 22

FIGURE 22

2701 TGGCGGGAAGCGGTGAGACGATGCGCGTACTGCTGACCTGCTCGCCAACGACACCCAC 2760
 A G R R * M R V L L T C F A N D T H
 OleG2

2761 TTCCACGGGCTGGTGCCGCTGGCGTGGCGCTGCCGGCCCGGACGAAGTCCGCGTG 2820
 F H G L V P L A W A L R A A G H E V R V

2821 GCCAGTCAGCCCGCCCTGTCCGACACGATCACCAAGCGGACTGACCGGGTGCCCGTG 2880
 A S Q P A L S D T I T Q A G L T A V P V

2881 GGCCGGGACACCGCCTTCCTGGAGCTGATGGGGAGATCGGGCGGACGTCCAGAAGTAC 2940
 G R D T A F L E L M G E I G A D V Q K Y

2941 TCCACCGGCATCGACCTGGGCGTCCGCGCGGAGCTGACGAGCTGGGAGTACCTGCTCGGC 3000
 S T G I D L G V R A E L T S W E Y L L G

3001 ATGCACACGACCCCTGGTGCCACGTTCTACTCGCTGGTCAACGACGAGCCGTTGTCGAC 3060
 M H T T L V P T F Y S L V N D E P F V D

3061 GGGCTCGTCGCGTGACCCCGGSCCTGGCGGCCCGACCTCATCTGTGGGAGCACTTCAGC 3120
 G L V A L T R A W R P D L I L W E H F S

3121 TTCGCCGGGCGTTGGCGGCGGGCCACCGGCACGCCCCACGCCCGCTGTGGGGG 3180
 F A G A L A A R A T G T P H A R V L W G

3181 TCGGACCTCATCGTCCGGTTCCGCCGGGACTTCCTCGCGGAGCGGCGAACC GCCCGCC 3240
 S D L I V R F R R D F L A E R A N R P A

3241 GAGCACCGGAGGACCCCATGGCGGAGTGGCTGGGCTGGCGGCCGGAACGGCTGGGCTCC 3300
 E H R E D P M A E W L G W A A E R L G S

3301 ACCTTCGACGAGGAGCTGGTGACCGGGCAGTGGACGATCGACCCCGCTGCCCGGAGCATG 3360
 T F D E E L V T G Q W T I D P L P R S M

FIGURE 22

56/60

3361 CCGTGCCCGGACGACGACGGTCCGATCGGTACGTCCGTACACGGCGGGCC 3420
R L P T G T T V P M R Y V P Y N G R A

3421 GTGGTCCCGCATGGGTCCGGCAGCGTGGCGGGCCCCGGATCTGCCTGACGCTCGGT 3480
V V P A W V R Q R A R R P R I C L T L G

3481 GTGTGGCCCGGACAGACCCCTGGGCGACGGCGTGTCTGGCGGAGGTGCTGGCCCGGCTG 3540
V S A R Q T L G D G V S L A E V L A A L

3541 GCGACGTGGACGCGGAGATCGTGGCCACGCTGGACGCCCTCCAGCGCAAGCTCCTGGGG 3600
G D V D A E I V A T L D A S Q R K L L G

3601 CCGGTGCCGACAAACGTCGGGCTGGTGGACTTCGTGCCCCCTGCACGCCCTGATGCCGACC 3660
P V P D N V R L V D F V P L H A L M P T

KpnI

3661 TGTTGGCGATCGTGACACCGGCGCGGTACCTGGCTGACGGCCGCGTCCACGGC 3720
C S A I V H H G G A G T W L T A A V H G

3721 GTCCCGCAGATCGTCTCGGTGACCTCTGGGACAACCTGCTGCGCCCGGCGAGACACAG 3780
V P Q I V L G D L W D N L L R A R Q T Q

3781 GCCGCGGCGCGGCTGTTTCATCCATCCGTCGAGGTACCGCGCGCGGCTCGGTGAG 3840
A A G A G L F I H P S E V T A A G L G E

3841 GCGTGCGCGCGGTGCTGACGGACCCCTCCATCCGGGCGCGCACAGCGCGTCCGGGAC 3900
G V R R V L T D P S I R A A A Q R V R D

3901 GAGATGAATGCAGACCGACGCGGCGGAGGTGCTACGGTGTGGAGCGGCTCGCCGCG 3960
E M N A E P T P G E V V T V L E R L A A

FIGURE 22

3961 AGCGGGGACGCGGACGAGGAGCGGGGAACCATCGGGCTGACACGGAGCCGACCAACCGG 4020
 S G G R G R G G G N H A G * M R A D T E P T T G
 oleM

BglII

4021 GTACGAGGACGAGTTCCGCGAGATCTACGACGCCGTGTACCGGGGGGGCAAGGACTA 4080
 Y E D E F A E I Y D A V Y R G R G K D Y

4081 CGCCGGGAGCGCAAGGACGTGGCGGACCTCGTGGGACCGGGTGCGGACGCGTCCTC 4140
 A G E A K D V A D L V R D R V P D A S S

4141 CCTCTGGACGTGGCCTGGGACGGGCGGCACCTGGGGACCTTCGCCACGCTCTTCGA 4200
 L L D V A C G T G A H L R H F A T L F D

4201 CGACGCGCGCGGTCTCGAACTGTCCGCGAGCATGCTGGACATCGCCGCTCCCGCATGCC 4260
 D A R G L E L S A S M L D I A R S R M P

4261 GGGCGTCCGCTGCACCAAGGGACATGCGATCCTTCGACCTGGGGCCACGCGTCTCCGC 4320
 G V P L H Q G G D M R S F D L G P R V S A

4321 GGTCACCTGCATGTTTCAGCTCCGTCCGGCCACCTGGCCACCACCGCGAACTCGACGCGAC 4380
 V T C M F S S V G H L A T T A E L D A T

4381 GCTGCGGTGCTTCGCGCGCACACCGCGCGCGCGGTGGCCGTCATCGAACCGTGGTG 4440
 L R C F A R H T R P G G V A V I E P W W

4441 GTTCCGGGAGACCTTCACCGACGGCTACGTGGCGGGTGACATCGTACGCGTCGACGGCGG 4500
 F P E T F T D G Y V A G D I V R V D G R

4501 GACCATCTCCCGGTGTCCCACTCGGTACGGGACGGCGCGCCACCCGATGGAGATCCA 4560
 T I S R V S H S V R D G G A T R M E I H

FIGURE 22

FIGURE 22

5221 CTGGCGCGCGCCACCAACGTGTGCTGCACGGCAGACCAACGAGACGCCCCACTG 5280
L A A R A T N V V L H A T T N E T P P L

5281 GACCGGCTGGCCCTGCGCTACGAGTCCGACAAAGTGGGGCGGCTCCACTGGTTCAACGGC 5340
D R L A L R Y E S D K W G G V H W F T G

BamHI

5341 CACTACGACCGGCACCTGCGGGCCGTGCGCGACCAAGCGGTGCGGATCCTGGAGATCGGC 5400
H Y D R H L R A V R D Q A V R I L E I G

5401 ATCGGGGCTACGACGACCTGTGCTGCCGAGCGGCCCTCACTGAAGATGTGGAAGCGGTAC 5460
I G G Y D D L L P S G A S L K M W K R Y

5461 TTCCCGCGCGGCTGTCTTCGSGCTGGACATCTTCGACAGTCGGCGTGCAGACCAAGCCGC 5520
F P R G L V F G V D I F D S R R A T S R

5521 GTGTCAAGACGCTCCGCGGCGCGGCGGAGGACGACCCGGAGTTCATGCGCCGTCGCCGAG 5580
V S R R S A A R Q D D P E F M R R V A E

5581 GAGCACGGGCGGTTTCGACGTCATCATCGACGAGCGGAGCCACATCAACGCACACATGCGG 5640
E H G P F D V I I D D G S H I N A H M R

5641 ACGTCGTTCTCGGTGATGTTCCCGCCACCTGCGCAACGGCGGCTTCTACGTCATCGAGGAC 5700
T S F S V M F P H L R N G G F Y V I E D

5701 ACCTTCACTCTACTGGCCCGGGTACGGAGGGCCATCCGGAGCCCGGTGCCCGTCCGGA 5760
T F T S Y W P G Y G G P S G A R C P S G

5761 ACAACCGCGCTGGAGATGGTCAAGGACTGATCGACTCGGTGCACTACGAGGAGCGGCCG 5820
T T A L E M V K G L I D S V H Y E E R P

FIGURE 22

5821 GACGGCGGGCCACGGCCGACTACATCGCCAGGAACCTCGTCGGGCTGCACGCCCTACCAA 5880
D G A A T A D Y I A R N L V G L H A Y Q
5881 ACGACCTCGTCTTCCTCGAGAAAGGGCGATCAACAAGGAGGGCGGCATCCCCCACCACCGTG 5940
T T S S S R R A I N K E G G I P H T V
5941 CCCCCGGAGCCGTTCTGGAAACGACAACCTAGCCACGGCCGCAACCCAGAGCCGGAAACCGCA 6000
P R E P F W N D N *
6001 CCACTGTCCGCGCCACCTCGGAACCACTCCAGCAAAGGACACACCGCTGTGACCGATAC 6060
6061 GCACACCGGACCGACACCGGGCCGACGGCGGTACC 6093
KpnI

FIGURE 22

LISTE DE SEQUENCES

(1) INFORMATIONS GENERALES:

5

(i) DEPOSANT:

(A) NOM: Hoechst Marion Roussel

(B) RUE: 1, Terrasse Bellini

(C) VILLE: PUTEAUX

10

(E) PAYS: FRANCE

(F) CODE POSTAL: 92800

(G) TELEPHONE: 01.49.91.57.27

(H) TELECOPIE: 01.49.91.46.10

15

(ii) TITRE DE L' INVENTION: Genes de biosynthese et de transfert des
6-desoxyhexoses chez Saccharopolyspora erythraea et chez
Streptomyces antibioticus et leur utilisation.

(iii) NOMBRE DE SEQUENCES: 61

20

(iv) FORME DECHIFFRABLE PAR ORDINATEUR:

(A) TYPE DE SUPPORT: Floppy disk

(B) ORDINATEUR: IBM PC compatible

(C) SYSTEME D' EXPLOITATION: PC-DOS/MS-DOS

25

(D) LOGICIEL: PatentIn Release #1.0, Version #1.30 (OEB)

(vi) DONNEES DE LA DEMANDE ANTERIEURE:

(A) NUMERO DE LA DEMANDE: FR 9709458

(B) DATE DE DEPOT: 25-JUL-1997

30

(vi) DONNEES DE LA DEMANDE ANTERIEURE:

(A) NUMERO DE LA DEMANDE: FR 9807411

(B) DATE DE DEPOT: 12-JUN-1998

35

(2) INFORMATIONS POUR LA SEQ ID NO: 1:

(i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 3439 paires de bases

40

(B) TYPE: nucléotide

(C) NOMBRE DE BRINS: double

(D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADNc

45

(vi) ORIGINE:

(A) ORGANISME: Saccharopolyspora erythraea

(ix) CARACTERISTIQUE:

- (A) NOM/CLE: CDS
5 (B) EMBLACEMENT: complement (48..1046)
(D) AUTRES INFORMATIONS: /function= "implique dans la
biosynthese du mycarose"
/gene= "eryBII"
- 10 (ix) CARACTERISTIQUE:
(A) NOM/CLE: CDS
(B) EMBLACEMENT: complement (2322..3404)
(D) AUTRES INFORMATIONS: /function= "implique dans la
biosynthese de la desosamine"
15 /gene= "eryCII"

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 1:

20 GCTTCACGCT CACCAGCCGT ATCCTTTCTC GGTTCCTCTT GTGCTCACTG CAACCAGGCT 60
TCCGGCGCCG CGCCGCCGGA GGCCACCGCG GGGAAGATCT CGTCCAGTTC GGACAGCGCC 120
25 TGCTCGTCCA GGGTCATCGC GGACGCCTTC AGCGCGGAGT CGAGCTGCTC GGGGGTTCGC 180
GGGCCGATGA CGGCGCCGGC GATGCCGGGC CGGGACAGCA CCCATGCGAG CCCCACCTCG 240
GCCGGGTCTT CGCCGAGGTT GCGGCAGAAC TTCTCGTAGG CCTCGATCGC CGGGCGCAGG 300
30 GACGGCAACA GCACCTGCGC ACGGCCCTGC GCCGACTTCA CCGCGGTGCC CGCGGCCAGC 360
TTCTCCAGCG CTCCGCTGAG CAGGCCGCCG TGCAGCGGCG ACCAGGCGAA GACGCCGAGC 420
35 CCGTAGGCCT GCGCGGCGGG CAGCACCTCC AGCTCGGCGT GCCGGACCGC CAGGTTGTAC 480
AGGCACTGGT GGGAGACCAT GCCCAGGGAG TGGCGGCGGG CGGCGTTCTC CTGCGCGGCG 540
GCGATGTGCC AGCCCGCGAA GTTCGACGAG CCGACGTAGG AGACCTTGCC GCTGGCGACG 600
40 AGGCTGTCCA TGGCCTGCCA CACCTCGTCC CACGGCGCGG ACCGGTCGAT GTGGTGCATC 660
TGGTAGACGT CGATGTGGTC GACGCCCAGC CTGCGCAGCG ATCCCTCGCA GGAGGCGATG 720
45 ATGTGCCGCG CCGACAGCCC GCTGTCGTTG ACGCGCTCGC TCATCTCGCC GCCGACCTTG 780

| | | | | | | | |
|----|------------|------------|------------|------------|------------|-------------|------|
| | GTCGCCAGCA | CGGTGTCCTC | GCGCCGTCCG | CCGCCCTGGG | CCAGCCACCT | GGCCACCAGC | 840 |
| | TCCTCGGTGT | GGCCCTTGTA | GAGCCGCCAG | CCGTACATGT | CGGCGGTGTC | GAGGCAGTTG | 900 |
| 5 | ATGCCGCGGT | CCCGGGCGTG | GTCCATCAGG | CGCAGCGCGT | CGTCGTCCTC | GACGCGTCCG | 960 |
| | CTGAAGTTCA | CCGTGCCGAG | CCAGAGCCTG | CTGGTGAGCA | GCGCGGAACG | CCCGAGCCGC | 1020 |
| | ACGTGCGTCG | CGGCGTCGGT | GGTCATCGTG | GTTCTCTCCT | TCCTGCGGCC | AGTTCTCTCGC | 1080 |
| 10 | AGATGCCGAC | GACCTCGGCC | GGTGACGGCT | CCGCGAGCAT | GTCGTCGCGC | ATCCGCGCCG | 1140 |
| | CGCCGCGCGC | GTGGGCCGGG | TCGTCGAGGA | CCCGCTTCAC | CGACTCCCGG | AGCTGGTCGG | 1200 |
| 15 | GGGTCAGCTC | GGGCACGGGC | AGCGCGATCC | CCGCCCCGAA | TTCCTGCGTG | CGCTGCGCGC | 1260 |
| | GCACGCCGGT | GTCCCAGCCG | TCGGGCAGGA | TCACCTGCGG | CACGCCGTGG | ATCGCCGCGG | 1320 |
| | TGTGCCAGCT | CCCGGGTCCG | CCGTGGTGCA | CCGTCGCCGC | GCAGGTCGGC | AGCAGCGCGT | 1380 |
| 20 | GCATCGGGAC | GAAGCCGACC | GTGCGGACGT | TGTCCGGGAT | GTTGCGGACG | CCTTCTAGCT | 1440 |
| | GCTGCGCGTC | GAAGGTCGCG | ATGATCTCGG | CGTCGACGTC | GCCGACGGCA | CCCAGCAGCT | 1500 |
| 25 | CCTCGATGGA | GACCTGCCCC | ATGCTGTTCT | CGCGGCTGGA | GATCCCGAGC | GTGAGGCACA | 1560 |
| | CGCGGCGGCG | CTCGGGCTCG | TCGTGCAGCC | ATTCCGGCAC | CACGGACGGC | CCGTTGTAGT | 1620 |
| | CGACGTAGCG | CATCCCGACG | GTCTTCAGGC | CGGTGTCGAG | CCTGATCGCG | GCCGGGGCGG | 1680 |
| 30 | GGTCGATCGT | CCACTGCCCC | ACGACCACCT | CCTCGTCGAA | GGCCGGGCCG | CCGTACTTCT | 1740 |
| | CCAGCGTCCA | GGTGAGCCAC | TCGGCGAGCG | GGTCCTCCCG | GTGCTCCTCC | GGCTGGTCGG | 1800 |
| 35 | GCAGCAGGCC | GAGGAAGTTC | TGCCGCGCCC | GGGTGGTGAT | GTCGGGTCCC | CACAGCAGCC | 1860 |
| | GCGCGTGCGG | CGTTCCGGTC | ACCGCCGCCG | CGATGGGCGC | GGCGAAGGTG | AGCGGCTCCC | 1920 |
| | AGATGACCAG | GTCGGGCCGC | CACTTCCGGC | AGAACGAGAC | CATGCCTTCG | ATGAGCGTGT | 1980 |
| 40 | CCGGGCTCAT | CAGGGCGTAG | AAGGTCGGGG | TGAGCACGGT | CTGCATGCCC | AGCAGGTGCT | 2040 |
| | CCCAGGTCAA | GGTGGCGGGG | TCCCCTCGC | TGAAGTCCAG | GCTCCGGACG | TAGTCGATGA | 2100 |
| 45 | TGTCGTGGCC | CGCGTGGGTC | ATGAAGTCCA | CGAGGTCGAC | GTCGGTGCCG | ACCGGGACGG | 2160 |

| | | | | | | | |
|----|------------|------------|------------|------------|------------|------------|------|
| | CGGTCAGCCC | GGCCGCGGTG | ATGTCCTCGG | TGAGCGCCGG | GGACGCGACC | ACGCGGACCT | 2220 |
| | CGTGCCCCGC | CGCGCGGAAC | GCCCATGCGA | GGGGGACGAG | GCCGAAGAGG | TGGCTCTTGC | 2280 |
| 5 | TGGCCATGGA | GGAGAAGACG | ACGCGCATCG | CGGTTACCTC | AGAGCTCGAC | GGGGCAGCGG | 2340 |
| | TTGGTTCCCC | GCAGGACGGG | TGATCGGCCG | CGCCGGACGA | CCGGGCCGCT | GGGCGTGAGT | 2400 |
| | CCGGGCAGCG | CCTTGGCCGC | GGCCCGCAGT | GCGGCGGTGG | CGAGCGCGGT | GACCAGCTCC | 2460 |
| 10 | TCCAGCCTGC | CGGGGTGGCC | GCGATGTGCC | GACAGCGCGC | GGTCGGCGTC | GGGGCGGTCC | 2520 |
| | ACGTCGAGGC | GGTCGGGCTC | GGCGAAGACC | TCCGGGTCGC | GGTTGGCCGC | CGCGACGACG | 2580 |
| 15 | ACCACGACCT | CCTCGCCTTC | GCCGATCACG | TGCTCGCCGA | GCCGCACCTC | TGCGGTGGCC | 2640 |
| | GTGCGCCGCT | CCAGGTGCAA | TGCCGGGTGC | AGGCGCAGCA | CCTCGGCGAC | GGTTCGCTGC | 2700 |
| | GCGGCGGCGG | GGTCGTCGGC | GATCCGTTCG | GCCAGCCCCG | GTTGCGCCGA | GACGGCCAGG | 2760 |
| 20 | ACCGCGTCGA | CCACGGTGTT | CGCGGTCATC | TCGGCCCCGG | CGAACAGGGC | GCGCAGTGCG | 2820 |
| | GGGTCGGCGG | GCAGTGCCGC | GACCGCTGCT | TCGGTCACCG | CGAGCTGCTG | CGGGCTGAGC | 2880 |
| 25 | TGGGCGTCCA | GGCTGACGCG | GGCGTCCCAC | GCGGCGCCGC | GCAGCACTCC | GGCTGCGCCG | 2940 |
| | AGCACGGCGG | TCATGCCCTG | CACCGGTACC | TGCCAGGCGA | AGTCGCCGAC | CAGGTCCAGC | 3000 |
| | CGCGCGCCCC | CGCCGGGGAG | CAGACCGGCG | AAGCTCTCCG | CCAGTTCCCC | GACGTCGGGG | 3060 |
| 30 | ACCTCGCCTT | CCCAGGACGC | GGCGTGACG | TCCCGGAACG | GCTGGGCCCA | CTCGGCGGGT | 3120 |
| | GGCGCGCCCC | CGGCCCCGAT | CCATTCCGGT | GTGCGTCCGG | TGGCGCGGGT | GAACGCGGGG | 3180 |
| 35 | TCGTCGAGCA | CCTGCCGGGC | GGTGGCGTGG | TCGGCCACCA | CCCACGTCTC | GGTGCGGCTG | 3240 |
| | CGCCGCACAC | CGGACTCGCG | CATCGAGCGG | TACCGGCGCT | GCGGGTCGTC | GTCGTGTCCG | 3300 |
| | CACAGCAGCA | TCGGGTAAGG | GTCGCCGTTG | CTGCCGTAAC | CCCAGTGCA | GCCGCGGATC | 3360 |
| 40 | ATCTGGAGCT | GCCTGCCCAG | CCCGGCGCGA | TCGGTCGTGG | TCATGAATTC | CCTCCGCCCA | 3420 |
| | GCCAGGCGTC | GATGTGCCG | | | | | 3439 |

45

(2) INFORMATIONS POUR LA SEQ ID NO: 2:

5

(i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 333 acides aminés

(B) TYPE: acide aminé

(D) CONFIGURATION: linéaire

5

(ii) TYPE DE MOLECULE: protéine

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 2:

```

Met Thr Thr Asp Ala Ala Thr His Val Arg Leu Gly Arg Ser Ala Leu
10   1           5           10           15

Leu Thr Ser Arg Leu Trp Leu Gly Thr Val Asn Phe Ser Gly Arg Val
           20           25           30

15 Glu Asp Asp Asp Ala Leu Arg Leu Met Asp His Ala Arg Asp Arg Gly
           35           40           45

Ile Asn Cys Leu Asp Thr Ala Asp Met Tyr Gly Trp Arg Leu Tyr Lys
           50           55           60

20 Gly His Thr Glu Glu Leu Val Gly Arg Trp Leu Ala Gln Gly Gly Gly
           65           70           75           80

Arg Arg Glu Asp Thr Val Leu Ala Thr Lys Val Gly Gly Glu Met Ser
25           85           90           95

Glu Arg Val Asn Asp Ser Gly Leu Ser Ala Arg His Ile Ile Ala Ser
           100          105          110

30 Cys Glu Gly Ser Leu Arg Arg Leu Gly Val Asp His Ile Asp Val Tyr
           115          120          125

Gln Met His His Ile Asp Arg Ser Ala Pro Trp Asp Glu Val Trp Gln
           130          135          140

35 Ala Met Asp Ser Leu Val Ala Ser Gly Lys Val Ser Tyr Val Gly Ser
           145          150          155          160

Ser Asn Phe Ala Gly Trp His Ile Ala Ala Ala Gln Glu Asn Ala Ala
40           165          170          175

Arg Arg His Ser Leu Gly Met Val Ser His Gln Cys Leu Tyr Asn Leu
           180          185          190

45 Ala Val Arg His Ala Glu Leu Glu Val Leu Pro Ala Ala Gln Ala Tyr
           195          200          205

```

Gly Leu Gly Val Phe Ala Trp Ser Pro Leu His Gly Gly Leu Leu Ser
 210 215 220

Gly Ala Leu Glu Lys Leu Ala Ala Gly Thr Ala Val Lys Ser Ala Gln
 5 225 230 235 240

Gly Arg Ala Gln Val Leu Leu Pro Ser Leu Arg Pro Ala Ile Glu Ala
 245 250 255

10 Tyr Glu Lys Phe Cys Arg Asn Leu Gly Glu Asp Pro Ala Glu Val Gly
 260 265 270

Leu Ala Trp Val Leu Ser Arg Pro Gly Ile Ala Gly Ala Val Ile Gly
 275 280 285

15 Pro Arg Thr Pro Glu Gln Leu Asp Ser Ala Leu Lys Ala Ser Ala Met
 290 295 300

Thr Leu Asp Glu Gln Ala Leu Ser Glu Leu Asp Glu Ile Phe Pro Ala
 20 305 310 315 320

Val Ala Ser Gly Gly Ala Ala Pro Glu Ala Trp Leu Gln
 325 330

25

(2) INFORMATIONS POUR LA SEQ ID NO: 3:

(i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 361 acides aminés
 30 (B) TYPE: acide aminé
 (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: protéine

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 3:

35

Met Thr Thr Thr Asp Arg Ala Gly Leu Gly Arg Gln Leu Gln Met Ile
 1 5 10 15

Arg Gly Leu His Trp Gly Tyr Gly Ser Asn Gly Asp Pro Tyr Pro Met
 40 20 25 30

Leu Leu Cys Gly His Asp Asp Asp Pro Gln Arg Arg Tyr Arg Ser Met
 35 40 45

45 Arg Glu Ser Gly Val Arg Arg Ser Arg Thr Glu Thr Trp Val Val Ala
 50 55 60

Asp His Ala Thr Ala Arg Gln Val Leu Asp Asp Pro Ala Phe Thr Arg
 65 70 75 80

Ala Thr Gly Arg Thr Pro Glu Trp Met Arg Ala Ala Gly Ala Pro Pro
 5 85 90 95

Ala Glu Trp Ala Gln Pro Phe Arg Asp Val His Ala Ala Ser Trp Glu
 100 105 110

10 Gly Glu Val Pro Asp Val Gly Glu Leu Ala Glu Ser Phe Ala Gly Leu
 115 120 125

Leu Pro Gly Ala Gly Ala Arg Leu Asp Leu Val Gly Asp Phe Ala Trp
 130 135 140

15 Gln Val Pro Val Gln Gly Met Thr Ala Val Leu Gly Ala Ala Gly Val
 145 150 155 160

Leu Arg Gly Ala Ala Trp Asp Ala Arg Val Ser Leu Asp Ala Gln Leu
 20 165 170 175

Ser Pro Gln Gln Leu Ala Val Thr Glu Ala Ala Val Ala Ala Leu Pro
 180 185 190

25 Ala Asp Pro Ala Leu Arg Ala Leu Phe Ala Gly Ala Glu Met Thr Ala
 195 200 205

Asn Thr Val Val Asp Ala Val Leu Ala Val Ser Ala Glu Pro Gly Leu
 210 215 220

30 Ala Glu Arg Ile Ala Asp Asp Pro Ala Ala Ala Gln Arg Thr Val Ala
 225 230 235 240

Glu Val Leu Arg Leu His Pro Ala Leu His Leu Glu Arg Arg Thr Ala
 35 245 250 255

Thr Ala Glu Val Arg Leu Gly Glu His Val Ile Gly Glu Gly Glu Glu
 260 265 270

40 Val Val Val Val Val Ala Ala Ala Asn Arg Asp Pro Glu Val Phe Ala
 275 280 285

Glu Pro Asp Arg Leu Asp Val Asp Arg Pro Asp Ala Asp Arg Ala Leu
 290 295 300

45

Ser Ala His Arg Gly His Pro Gly Arg Leu Glu Glu Leu Val Thr Ala
305 310 315 320

Leu Ala Thr Ala Ala Leu Arg Ala Ala Ala Lys Ala Leu Pro Gly Leu
5 325 330 335

Thr Pro Ser Gly Pro Val Val Arg Arg Arg Arg Ser Pro Val Leu Arg
340 345 350

10 Gly Thr Asn Arg Cys Pro Val Glu Leu
355 360

(2) INFORMATIONS POUR LA SEQ ID NO: 4:

- 15 (i) CARACTERISTIQUES DE LA SEQUENCE:
(A) LONGUEUR: 1266 paires de bases
(B) TYPE: nucléotide
(C) NOMBRE DE BRINS: double
(D) CONFIGURATION: linéaire
- 20 (ii) TYPE DE MOLECULE: ADNc
- (vi) ORIGINE:
(A) ORGANISME: Saccharopolyspora erythraea
- 25 (ix) CARACTERISTIQUE:
(A) NOM/CLE: CDS
(B) EMBLACEMENT: complement (4..1266)
(D) AUTRES INFORMATIONS: /function= "implique dans la
30 biosynthese de la desosamine"
/gene= "eryCIII"
/note= "SEQ ID NO 1 DE 1046 A 2308"

35 (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 4:

TCATCGTGGT TCTCTCCTTC CTGCGGCCAG TTCCTCGCAG ATGCCGACGA CCTCGGCCGG 60

TGACGGCTCC GCGAGCATGT CGTCGCGCAT CCGCGCCGCG CCGGCGCGGT GGGCCGGGTC 120

40 GTCGAGGACC CGCTTCACCG ACTCCCGGAG CTGGTCGGGG GTCAGCTCGG GCACGGGCAG 180

CGCGATCCCC GCCCCGAATT CCTGCGTGCG CTGCGCGCGC ACGCCGGTGT CCCAGCCGTC 240

45 GGGCAGGATC ACCTGCGGCA CGCCGTGGAT CGCCGCGGTG TGCCAGCTCC CGGGTCCGCC 300

GTGGTGCACC GTCGCCGCGC AGGTCGGCAG CAGCGCGTGC ATCGGGACGA AGCCGACCGT 360
GCGGACGTTG TCCGGGATGT TCGCGACGCC TTCTAGCTGC TCGCGCTCGA AGGTCGCGAT 420
5 GATCTCGGCG TCGACGTGCG CGACGGCACC CAGCAGCTCC TCGATGGAGA CCTGCCCCGAT 480
GCTGTTCTCG CGGCTGGAGA TCCCGAGCGT GAGGCACACG CGGCGGCGCT CGGGCTCGTC 540
GTGCAGCCAT TCCGGCACCA CGGACGGCCC GTTGTAGTCG ACGTAGCGCA TCCCGACGGT 600
10 CTTCAGGCCG GTGTCGAGCC TGATCGCGGC CGGGGCGGGG TCGATCGTCC ACTGCCCGAC 660
GACCACCTCC TCGTCGAAGG CCGGGCCGCC GTACTTCTCC AGCGTCCAGG TGAGCCACTC 720
15 GCGGAGCGGG TCCTCCCGGT GCTCCTCCGG CTGGTCGGGC AGCAGGCCGA GGAAGTTCTG 780
CCGCGCCCCG GTGGTGATGT CGGGTCCCCA CAGCAGCCGC GCGTGGGCG TTCCGGTCAC 840
CGCCGCCGCG ATGGGCGCGG CGAAGGTGAG CGGCTCCCAG ATGACCAGGT CGGGCCGCCA 900
20 CTTCGGCAG AACGAGACCA TGCCTTCGAT GAGCGTGTCC GGGCTCATCA GGGCGTAGAA 960
GGTCGGGGTG AGCACGGTCT GCATGCCAG CAGGTGCTCC CAGGTCAAGG TGCGGGGGTC 1020
25 CCGCTCGCTG AAGTCCAGGC TCCGGACGTA GTCGATGATG TCGTGGCCCG CGTGGGTCAT 1080
GAAGTCCACG AGGTCGACGT CGGTGCCGAC CGGACGGCG GTCAGCCCGG CCGCGGTGAT 1140
GTCCTCGGTG AGCGCCGGGG ACGCGACCAC GCGGACCTCG TGCCCCGCG CGCGGAACGC 1200
30 CCATGCGAGG GGGACGAGGC CGAAGAGGTG GCTCTTGCTG GCCATGGAGG AGAAGACGAC 1260
GCGCAT 1266

35

(2) INFORMATIONS POUR LA SEQ ID NO: 5:

(i) CARACTERISTIQUES DE LA SEQUENCE:

- 40 (A) LONGUEUR: 421 acides aminés
(B) TYPE: acide aminé
(D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: protéine

45 (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 5:

10

Met Arg Val Val Phe Ser Ser Met Ala Ser Lys Ser His Leu Phe Gly
 1 5 10 15
 Leu Val Pro Leu Ala Trp Ala Phe Arg Ala Ala Gly His Glu Val Arg
 5 20 25 30
 Val Val Ala Ser Pro Ala Leu Thr Glu Asp Ile Thr Ala Ala Gly Leu
 35 40 45
 10 Thr Ala Val Pro Val Gly Thr Asp Val Asp Leu Val Asp Phe Met Thr
 50 55 60
 His Ala Gly His Asp Ile Ile Asp Tyr Val Arg Ser Leu Asp Phe Ser
 65 70 75 80
 15 Glu Arg Asp Pro Ala Thr Leu Thr Trp Glu His Leu Leu Gly Met Gln
 85 90 95
 Thr Val Leu Thr Pro Thr Phe Tyr Ala Leu Met Ser Pro Asp Thr Leu
 20 100 105 110
 Ile Glu Gly Met Val Ser Phe Cys Arg Lys Trp Arg Pro Asp Leu Val
 115 120 125
 25 Ile Trp Glu Pro Leu Thr Phe Ala Ala Pro Ile Ala Ala Ala Val Thr
 130 135 140
 Gly Thr Pro His Ala Arg Leu Leu Trp Gly Pro Asp Ile Thr Thr Arg
 145 150 155 160
 30 Ala Arg Gln Asn Phe Leu Gly Leu Leu Pro Asp Gln Pro Glu Glu His
 165 170 175
 Arg Glu Asp Pro Leu Ala Glu Trp Leu Thr Trp Thr Leu Glu Lys Tyr
 35 180 185 190
 Gly Gly Pro Ala Phe Asp Glu Glu Val Val Val Gly Gln Trp Thr Ile
 195 200 205
 40 Asp Pro Ala Pro Ala Ala Ile Arg Leu Asp Thr Gly Leu Lys Thr Val
 210 215 220
 Gly Met Arg Tyr Val Asp Tyr Asn Gly Pro Ser Val Val Pro Glu Trp
 225 230 235 240
 45

11

Leu His Asp Glu Pro Glu Arg Arg Arg Val Cys Leu Thr Leu Gly Ile
 245 250 255

Ser Ser Arg Glu Asn Ser Ile Gly Gln Val Ser Ile Glu Glu Leu Leu
 5 260 265 270

Gly Ala Val Gly Asp Val Asp Ala Glu Ile Ile Ala Thr Phe Asp Ala
 275 280 285

10 Gln Gln Leu Glu Gly Val Ala Asn Ile Pro Asp Asn Val Arg Thr Val
 290 295 300

Gly Phe Val Pro Met His Ala Leu Leu Pro Thr Cys Ala Ala Thr Val
 305 310 315 320

15 His His Gly Gly Pro Gly Ser Trp His Thr Ala Ala Ile His Gly Val
 325 330 335

Pro Gln Val Ile Leu Pro Asp Gly Trp Asp Thr Gly Val Arg Ala Gln
 20 340 345 350

Arg Thr Gln Glu Phe Gly Ala Gly Ile Ala Leu Pro Val Pro Glu Leu
 355 360 365

25 Thr Pro Asp Gln Leu Arg Glu Ser Val Lys Arg Val Leu Asp Asp Pro
 370 375 380

Ala His Arg Ala Gly Ala Ala Arg Met Arg Asp Asp Met Leu Ala Glu
 385 390 395 400

30 Pro Ser Pro Ala Glu Val Val Gly Ile Cys Glu Glu Leu Ala Ala Gly
 405 410 415

Arg Arg Glu Pro Arg
 35 420

(2) INFORMATIONS POUR LA SEQ ID NO: 6:

(i) CARACTERISTIQUES DE LA SEQUENCE:

- 40 (A) LONGUEUR: 8160 paires de bases
- (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: double
- (D) CONFIGURATION: linéaire

45 (ii) TYPE DE MOLECULE: ADN (génomique)

- (vi) ORIGINE:
 (A) ORGANISME: Saccharopolyspora erythraea
- (ix) CARACTERISTIQUE:
5 (A) NOM/CLE: CDS
 (B) EMLACEMENT:242..1207
 (D) AUTRES INFORMATIONS:/function= "implique dans la
 biosynthese du mycarose"
 /gene= "eryBIV"
10 /transl_except= (pos: 242 .. 244, aa: Met)
- (ix) CARACTERISTIQUE:
 (A) NOM/CLE: CDS
 (B) EMLACEMENT:1210..2454
15 (D) AUTRES INFORMATIONS:/function= "implique dans la
 biosynthese du mycarose"
 /gene= "eryBV"
 /transl_except= (pos: 1210 .. 1212, aa: Met)
- (ix) CARACTERISTIQUE:
20 (A) NOM/CLE: CDS
 (B) EMLACEMENT:2510..3220
 (D) AUTRES INFORMATIONS:/function= "implique dans la
 biosynthese de la desosamine"
25 /gene= "eryCVI"
- (ix) CARACTERISTIQUE:
 (A) NOM/CLE: CDS
 (B) EMLACEMENT:3308..4837
30 (D) AUTRES INFORMATIONS:/function= "implique dans la
 biosynthese du mycarose"
 /gene= "eryBVI"
 /transl_except= (pos: 3308 .. 3310, aa: Met)
- (ix) CARACTERISTIQUE:
35 (A) NOM/CLE: CDS
 (B) EMLACEMENT:6080..7546
 (D) AUTRES INFORMATIONS:/function= "implique dans la
 biosynthese de la desosamine"
40 /gene= "eryCV"
- (ix) CARACTERISTIQUE:
 (A) NOM/CLE: CDS
 (B) EMLACEMENT:7578..8156
45 (D) AUTRES INFORMATIONS:/function= "implique dans la
 biosynthese du mycarose"

/gene= "eryBVII"

/transl_except= (pos: 7578 .. 7580, aa: Met)

(ix) CARACTERISTIQUE:

- 5 (A) NOM/CLE: mat_peptide
(B) EMPLACEMENT:242

10 (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 6:

```

TTTGACAGGT CCGCCACGCG TCCCCCTACT CGACGACCAC GCAATGGGCG AACAAATATAG      60
GAAGGATCAA GAGGTTGACA TCGCCTCGTC GAGCCAACGA ACCTGTGAAC ATCTGCATGT      120
15 TGACAAGATC AACGGCGGCT ACCTACTGTG GTGGCCCAGT GACGGGTTGC CGCACATCGC      180
GCTGGGGAGA TTCTTTGAAT TTCGCCCCTA GCACCGACCT GGAAAGCGAG CAAATGCTCC      240

20 G GTG AAT GGG ATC AGT GAT TCC CCG CGT CAA TTG ATC ACC CTT CTG      286
   Met Asn Gly Ile Ser Asp Ser Pro Arg Gln Leu Ile Thr Leu Leu
     1           5           10           15

   GGC GCT TCC GGC TTC GTC GGG AGC GCG GTT CTG CGC GAG CTG CGC GAC      334
25 Gly Ala Ser Gly Phe Val Gly Ser Ala Val Leu Arg Glu Leu Arg Asp
     20           25           30

   CAC CCG GTC CGG CTG CGC GCG GTG TCC CGC GGC GGA GCG CCC GCG GTT      382
   His Pro Val Arg Leu Arg Ala Val Ser Arg Gly Gly Ala Pro Ala Val
30           35           40           45

   CCG CCC GGC GCC GCG GAG GTC GAG GAC CTG CGC GCC GAC CTG CTG GAA      430
   Pro Pro Gly Ala Ala Glu Val Glu Asp Leu Arg Ala Asp Leu Leu Glu
     50           55           60

35 CCG GGC CGG GCC GCC GCC GCG ATC GAG GAC GCC GAC GTG ATC GTG CAC      478
   Pro Gly Arg Ala Ala Ala Ala Ile Glu Asp Ala Asp Val Ile Val His
     65           70           75

40 CTG GTG GCG CAC GCA GCG GGC GGT TCC ACC TGG CGC AGC GCC ACC TCC      526
   Leu Val Ala His Ala Ala Gly Gly Ser Thr Trp Arg Ser Ala Thr Ser
     80           85           90           95

   GAC CCG GAA GCC GAG CGG GTC AAC GTC GGC CTG ATG CAC GAC CTC GTC      574
45 Asp Pro Glu Ala Glu Arg Val Asn Val Gly Leu Met His Asp Leu Val
     100           105           110

```

14

| | | |
|----|---|------|
| | GGC GCG CTG CAC GAT CGC CGC AGG TCG ACG CCG CCC GTG TTG CTC TAC | 622 |
| | Gly Ala Leu His Asp Arg Arg Arg Ser Thr Pro Pro Val Leu Leu Tyr | |
| | 115 120 125 | |
| 5 | GCG AGC ACC GCA CAG GCC GCG AAC CCG TCG GCG GCC AGC AGG TAC GCG | 670 |
| | Ala Ser Thr Ala Gln Ala Ala Asn Pro Ser Ala Ala Ser Arg Tyr Ala | |
| | 130 135 140 | |
| | CAG CAG AAG ACC GAG GCC GAG CGC ATC CTG CGC AAA GCC ACC GAC GAG | 718 |
| 10 | Gln Gln Lys Thr Glu Ala Glu Arg Ile Leu Arg Lys Ala Thr Asp Glu | |
| | 145 150 155 | |
| | GGC CGG GTG CGC GGC GTG ATC CTG CGG CTG CCC GCC GTC TAC GGC CAG | 766 |
| | Gly Arg Val Arg Gly Val Ile Leu Arg Leu Pro Ala Val Tyr Gly Gln | |
| 15 | 160 165 170 175 | |
| | AGC GGC CCG TCC GGC CCC ATG GGG CGG GGC GTG GTC GCA GCG ATG ATC | 814 |
| | Ser Gly Pro Ser Gly Pro Met Gly Arg Gly Val Val Ala Ala Met Ile | |
| | 180 185 190 | |
| 20 | CGG CGT GCC CTC GCC GGC GAG CCG CTC ACC ATG TGG CAC GAC GGC GGC | 862 |
| | Arg Arg Ala Leu Ala Gly Glu Pro Leu Thr Met Trp His Asp Gly Gly | |
| | 195 200 205 | |
| 25 | GTG CGC CGC GAC CTG CTG CAC GTC GAG GAC GTG GCC ACC GCG TTC GCC | 910 |
| | Val Arg Arg Asp Leu Leu His Val Glu Asp Val Ala Thr Ala Phe Ala | |
| | 210 215 220 | |
| | GCC GCG CTG GAG CAC CAC GAC GCG CTG GCC GGC GGC ACG TGG GCG CTG | 958 |
| 30 | Ala Ala Leu Glu His His Asp Ala Leu Ala Gly Gly Thr Trp Ala Leu | |
| | 225 230 235 | |
| | GGC GCC GAC CGA TCC GAG CCG CTC GGC GAC ATC TTC CGG GCC GTC TCC | 1006 |
| | Gly Ala Asp Arg Ser Glu Pro Leu Gly Asp Ile Phe Arg Ala Val Ser | |
| 35 | 240 245 250 255 | |
| | GGC AGC GTC GCC CGG CAG ACC GGC AGC CCC GCC GTC GAC GTG GTC ACC | 1054 |
| | Gly Ser Val Ala Arg Gln Thr Gly Ser Pro Ala Val Asp Val Val Thr | |
| | 260 265 270 | |
| 40 | GTG CCC GCG CCC GAG CAC GCC GAG GCC AAC GAC TTC CGC AGC GAC GAC | 1102 |
| | Val Pro Ala Pro Glu His Ala Glu Ala Asn Asp Phe Arg Ser Asp Asp | |
| | 275 280 285 | |

45

15

| | | |
|----|---|------|
| | ATC GAC TCC ACC GAG TTC CGC AGC CGG ACC GGC TGG CGC CCC CGG GTT | 1150 |
| | Ile Asp Ser Thr Glu Phe Arg Ser Arg Thr Gly Trp Arg Pro Arg Val | |
| | 290 295 300 | |
| 5 | TCC CTC ACC GAC GGC ATC GAC CGG ACG GTG GCC GCC CTG ACC CCC ACC | 1198 |
| | Ser Leu Thr Asp Gly Ile Asp Arg Thr Val Ala Ala Leu Thr Pro Thr | |
| | 305 310 315 | |
| | GAG GAG CAC TA GTG CGG GTA CTG CTG ACG TCC TTC GCG CAC CGC ACG | 1245 |
| 10 | Glu Glu His Met Arg Val Leu Leu Thr Ser Phe Ala His Arg Thr | |
| | 320 1 5 10 | |
| | CAC TTC CAG GGA CTG GTC CCG CTG GCG TGG GCG CTG CGC ACC GCG GGT | 1293 |
| | His Phe Gln Gly Leu Val Pro Leu Ala Trp Ala Leu Arg Thr Ala Gly | |
| 15 | 15 20 25 | |
| | CAC GAC GTG CGC GTG GCC GCC CAG CCC GCG CTC ACC GAC GCG GTC ATC | 1341 |
| | His Asp Val Arg Val Ala Ala Gln Pro Ala Leu Thr Asp Ala Val Ile | |
| | 30 35 40 | |
| 20 | | |
| | GGC GCC GGT CTC ACC GCG GTA CCC GTC GGC TCC GAC CAC CGG CTG TTC | 1389 |
| | Gly Ala Gly Leu Thr Ala Val Pro Val Gly Ser Asp His Arg Leu Phe | |
| | 45 50 55 60 | |
| 25 | GAC ATC GTC CCG GAA GTC GCC GCT CAG GTG CAC CGC TAC TCC TTC TAC | 1437 |
| | Asp Ile Val Pro Glu Val Ala Ala Gln Val His Arg Tyr Ser Phe Tyr | |
| | 65 70 75 | |
| | CTG GAC TTC TAC CAC CGC GAG CAG GAG CTG CAC TCG TGG GAG TTC CTG | 1485 |
| 30 | Leu Asp Phe Tyr His Arg Glu Gln Glu Leu His Ser Trp Glu Phe Leu | |
| | 80 85 90 | |
| | CTC GGC ATG CAG GAG GCC ACC TCG CGG TGG GTA TAC CCG GTG GTC AAC | 1533 |
| | Leu Gly Met Gln Glu Ala Thr Ser Arg Trp Val Tyr Pro Val Val Asn | |
| 35 | 95 100 105 | |
| | AAC GAC TCC TTC GTC GCC GAG CTG GTC GAC TTC GCC CGG GAC TGG CGT | 1581 |
| | Asn Asp Ser Phe Val Ala Glu Leu Val Asp Phe Ala Arg Asp Trp Arg | |
| | 110 115 120 | |
| 40 | | |
| | CCT GAC CTG GTG CTC TGG GAG CCG TTC ACC TTC GCC GGC GCC GTC GCG | 1629 |
| | Pro Asp Leu Val Leu Trp Glu Pro Phe Thr Phe Ala Gly Ala Val Ala | |
| | 125 130 135 140 | |
| 45 | | |

16

| | | |
|----|---|------|
| | GCC CGG GCC TGC GGA GCC GCG CAC GCC CGG CTG CTG TGG GGC AGC GAC | 1677 |
| | Ala Arg Ala Cys Gly Ala Ala His Ala Arg Leu Leu Trp Gly Ser Asp | |
| | 145 150 155 | |
| 5 | CTC ACC GGC TAC TTC CGC GGC CGG TTC CAG GCG CAA CGC CTG CGA CGG | 1725 |
| | Leu Thr Gly Tyr Phe Arg Gly Arg Phe Gln Ala Gln Arg Leu Arg Arg | |
| | 160 165 170 | |
| | CCG CCG GAG GAC CGG CCG GAC CCG CTG GGC ACG TGG CTG ACC GAG GTC | 1773 |
| 10 | Pro Pro Glu Asp Arg Pro Asp Pro Leu Gly Thr Trp Leu Thr Glu Val | |
| | 175 180 185 | |
| | GCG GGG CGC TTC GGC GTC GAA TTC GGC GAG GAC CTC GCG GTC GGG CAG | 1821 |
| | Ala Gly Arg Phe Gly Val Glu Phe Gly Glu Asp Leu Ala Val Gly Gln | |
| 15 | 190 195 200 | |
| | TGG TCG GTC GAC CAG TTG CCG CCG AGT TTC CGG CTG GAC ACC GGA ATG | 1869 |
| | Trp Ser Val Asp Gln Leu Pro Pro Ser Phe Arg Leu Asp Thr Gly Met | |
| | 205 210 215 220 | |
| 20 | GAA ACC GTT GTC GCG CGG ACC CTG CCC TAC AAC GGC GCG TCG GTG GTT | 1917 |
| | Glu Thr Val Val Ala Arg Thr Leu Pro Tyr Asn Gly Ala Ser Val Val | |
| | 225 230 235 | |
| 25 | CCG GAC TGG CTC AAG AAG GGC AGT GCG ACT CGA CGC ATC TGC ATT ACC | 1965 |
| | Pro Asp Trp Leu Lys Lys Gly Ser Ala Thr Arg Arg Ile Cys Ile Thr | |
| | 240 245 250 | |
| | GGA GGG TTC TCC GGA CTC GGG CTC GCC GCC GAT GCC GAT CAG TTC GCG | 2013 |
| 30 | Gly Gly Phe Ser Gly Leu Gly Leu Ala Ala Asp Ala Asp Gln Phe Ala | |
| | 255 260 265 | |
| | CGG ACG CTC GCG CAG CTC GCG CGA TTC GAT GGC GAA ATC GTG GTT ACG | 2061 |
| | Arg Thr Leu Ala Gln Leu Ala Arg Phe Asp Gly Glu Ile Val Val Thr | |
| 35 | 270 275 280 | |
| | GGT TCC GGT CCG GAT ACC TCC GCG GTA CCG GAC AAC ATT CGT TTG GTG | 2109 |
| | Gly Ser Gly Pro Asp Thr Ser Ala Val Pro Asp Asn Ile Arg Leu Val | |
| | 285 290 295 300 | |
| 40 | GAT TTC GTT CCG ATG GGC GTT CTG CTC CAG AAC TGC GCG GCG ATC ATC | 2157 |
| | Asp Phe Val Pro Met Gly Val Leu Leu Gln Asn Cys Ala Ala Ile Ile | |
| | 305 310 315 | |

45

17

| | | |
|----|---|------|
| | CAC CAC GGC GGG GCC GGA ACC TGG GCC ACG GCA CTG CAC CAC GGA ATT | 2205 |
| | His His Gly Gly Ala Gly Thr Trp Ala Thr Ala Leu His His Gly Ile | |
| | 320 325 330 | |
| 5 | CCG CAA ATA TCA GTT GCA CAT GAA TGG GAT TGC ATG CTA CGC GGC CAG | 2253 |
| | Pro Gln Ile Ser Val Ala His Glu Trp Asp Cys Met Leu Arg Gly Gln | |
| | 335 340 345 | |
| | CAG ACC GCG GAA CTG GGC GCG GGA ATC TAC CTC CGG CCG GAC GAG GTC | 2301 |
| 10 | Gln Thr Ala Glu Leu Gly Ala Gly Ile Tyr Leu Arg Pro Asp Glu Val | |
| | 350 355 360 | |
| | GAT GCC GAC TCA TTG GCG AGC GCC CTC ACC CAG GTG GTC GAG GAC CCC | 2349 |
| | Asp Ala Asp Ser Leu Ala Ser Ala Leu Thr Gln Val Val Glu Asp Pro | |
| 15 | 365 370 375 380 | |
| | ACC TAC ACC GAG AAC GCG GTG AAG CTT CGC GAG GAG GCG CTG TCC GAC | 2397 |
| | Thr Tyr Thr Glu Asn Ala Val Lys Leu Arg Glu Glu Ala Leu Ser Asp | |
| | 385 390 395 | |
| 20 | | |
| | CCG ACG CCG CAG GAG ATC GTC CCG CGA CTG GAG GAA CTC ACG CGC CGC | 2445 |
| | Pro Thr Pro Gln Glu Ile Val Pro Arg Leu Glu Glu Leu Thr Arg Arg | |
| | 400 405 410 | |
| 25 | CAC GCC GGC TAGCGGTTTC CGACCGACAA GTCCGTCCGA CAGCACACCT | 2494 |
| | His Ala Gly | |
| | 415 | |
| | CCGGAGGGGAG CAGGG ATG TAC GAG GGC GGG TTC GCC GAG CTT TAC GAC CGG | 2545 |
| 30 | Met Tyr Glu Gly Gly Phe Ala Glu Leu Tyr Asp Arg | |
| | 1 5 10 | |
| | TTC TAC CGC GGC CGG GGC AAG GAC TAC GCG GCC GAG GCC GCG CAG GTC | 2593 |
| | Phe Tyr Arg Gly Arg Gly Lys Asp Tyr Ala Ala Glu Ala Ala Gln Val | |
| 35 | 15 20 25 | |
| | GCG CGG CTG GTC AGA GAC CGC CTG CCC TCG GCT TCC TCG CTG CTC GAC | 2641 |
| | Ala Arg Leu Val Arg Asp Arg Leu Pro Ser Ala Ser Ser Leu Leu Asp | |
| | 30 35 40 | |
| 40 | | |
| | GTG GCC TGC GGG ACC GGC ACC CAC CTG CGC CGG TTC GCC GAC CTC TTC | 2689 |
| | Val Ala Cys Gly Thr Gly Thr His Leu Arg Arg Phe Ala Asp Leu Phe | |
| | 45 50 55 60 | |
| 45 | | |

18

| | | |
|----|---|------|
| | GAC GAC GTG ACC GGG CTG GAG CTG TCG GCG GCG ATG ATC GAG GTC GCC | 2737 |
| | Asp Asp Val Thr Gly Leu Glu Leu Ser Ala Ala Met Ile Glu Val Ala | |
| | 65 70 75 | |
| 5 | CGG CCG CAG CTC GGC GGC ATC CCG GTG CTG CAG GGC GAC ATG CGC GAC | 2785 |
| | Arg Pro Gln Leu Gly Gly Ile Pro Val Leu Gln Gly Asp Met Arg Asp | |
| | 80 85 90 | |
| | TTC GCG CTG GAT CGC GAG TTC GAC GCC GTC ACC TGC ATG TTC AGC TCC | 2833 |
| 10 | Phe Ala Leu Asp Arg Glu Phe Asp Ala Val Thr Cys Met Phe Ser Ser | |
| | 95 100 105 | |
| | ATC GGG CAC ATG CGC GAC GGC GCC GAG CTG GAC CAG GCG CTG GCG TCC | 2881 |
| | Ile Gly His Met Arg Asp Gly Ala Glu Leu Asp Gln Ala Leu Ala Ser | |
| 15 | 110 115 120 | |
| | TTC GCC CGC CAC CTC GCC CCC GGC GGC GTC GTG GTG GTC GAA CCG TGG | 2929 |
| | Phe Ala Arg His Leu Ala Pro Gly Gly Val Val Val Val Glu Pro Trp | |
| | 125 130 135 140 | |
| 20 | TGG TTC CCG GAG GAC TTC CTC GAC GGC TAC GTG GCC GGT GAC GTG GTG | 2977 |
| | Trp Phe Pro Glu Asp Phe Leu Asp Gly Tyr Val Ala Gly Asp Val Val | |
| | 145 150 155 | |
| 25 | CGC GAC GGC GAC CTG ACG ATC TCG CGC GTC TCG CAC TCC GTG CGC GCC | 3025 |
| | Arg Asp Gly Asp Leu Thr Ile Ser Arg Val Ser His Ser Val Arg Ala | |
| | 160 165 170 | |
| | GGC GGC GCG ACC CGG ATG GAG ATC CAC TGG GTC GTG GCC GAC GCG GTG | 3073 |
| 30 | Gly Gly Ala Thr Arg Met Glu Ile His Trp Val Val Ala Asp Ala Val | |
| | 175 180 185 | |
| | AAC GGT CCG CGG CAC CAC GTG GAG CAC TAC GAG ATC ACG CTC TTC GAG | 3121 |
| | Asn Gly Pro Arg His His Val Glu His Tyr Glu Ile Thr Leu Phe Glu | |
| 35 | 190 195 200 | |
| | CGG CAG CAG TAC GAG AAG GCC TTC ACC GCG GCC GGT TGC GCT GTG CAG | 3169 |
| | Arg Gln Gln Tyr Glu Lys Ala Phe Thr Ala Ala Gly Cys Ala Val Gln | |
| | 205 210 215 220 | |
| 40 | TAC CTG GAG GGC GGA CCC TCC GGA CGC GGG TTG TTC GTC GGT GTG CGC | 3217 |
| | Tyr Leu Glu Gly Gly Pro Ser Gly Arg Gly Leu Phe Val Gly Val Arg | |
| | 225 230 235 | |
| 45 | GGA TGACCCGTGC GTTCGCGTTT TCCGTTCTCTG GCACAGGTGA TCCGCTCCAC | 3270 |
| | Gly | |

| | | |
|----|--|------|
| | GGGCCCTTTC CCCGCCGTGA CCGGACCCTT ACAGTGA GTG CGG GTC TTG ATC GAC | 3325 |
| | Met Arg Val Leu Ile Asp | |
| | 1 5 | |
| 5 | AAC GCC CGG CGG CAG CAA GCG GAG CCG TCG ACG ACA CCG CAG GGA GAG | 3373 |
| | Asn Ala Arg Arg Gln Gln Ala Glu Pro Ser Thr Thr Pro Gln Gly Glu | |
| | 10 15 20 | |
| 10 | TCG ATG GGT GAT CGG ACC GGC GAC CGG ACG ATT CCG GAA TCC TCG CAG | 3421 |
| | Ser Met Gly Asp Arg Thr Gly Asp Arg Thr Ile Pro Glu Ser Ser Gln | |
| | 25 30 35 | |
| | ACC GCA ACG CGT TTC CTG CTC GGC GAC GGC GGA ATC CCC ACC GCC ACG | 3469 |
| 15 | Thr Ala Thr Arg Phe Leu Leu Gly Asp Gly Gly Ile Pro Thr Ala Thr | |
| | 40 45 50 | |
| | GCG GAA ACC CAC GAC TGG CTG ACC CGC AAC GGC GCC GAG CAG CGG CTC | 3517 |
| | Ala Glu Thr His Asp Trp Leu Thr Arg Asn Gly Ala Glu Gln Arg Leu | |
| 20 | 55 60 65 70 | |
| | GAG GTG GCG CGC GTG CCG TTC AGC GCC ATG GAC CGC TGG TCG TTC CAG | 3565 |
| | Glu Val Ala Arg Val Pro Phe Ser Ala Met Asp Arg Trp Ser Phe Gln | |
| | 75 80 85 | |
| 25 | CCC GAG GAC GGC AGG CTC GCC CAC GAG TCC GGC CGC TTC TTC TCC ATC | 3613 |
| | Pro Glu Asp Gly Arg Leu Ala His Glu Ser Gly Arg Phe Phe Ser Ile | |
| | 90 95 100 | |
| 30 | GAG GGC CTG CAC GTG CGG ACG AAC TTC GGC TGG CGG CGG GAC TGG ATC | 3661 |
| | Glu Gly Leu His Val Arg Thr Asn Phe Gly Trp Arg Arg Asp Trp Ile | |
| | 105 110 115 | |
| | CAG CCC ATC ATC GTG CAG CCC GAG ATC GGC TTC CTC GGC CTC ATC GTC | 3709 |
| 35 | Gln Pro Ile Ile Val Gln Pro Glu Ile Gly Phe Leu Gly Leu Ile Val | |
| | 120 125 130 | |
| | AAG GAG TTC GAC GGT GTG CTG CAC GTG CTG GCG CAG GCC AAG GCC GAG | 3757 |
| | Lys Glu Phe Asp Gly Val Leu His Val Leu Ala Gln Ala Lys Ala Glu | |
| 40 | 135 140 145 150 | |
| | CCG GGC AAC ATC AAC GCC GTC CAG CTC TCC CCG ACC CTG CAG GCG ACC | 3805 |
| | Pro Gly Asn Ile Asn Ala Val Gln Leu Ser Pro Thr Leu Gln Ala Thr | |
| | 155 160 165 | |
| 45 | | |

20

| | | |
|----|---|------|
| | CGC AGC AAC TAC ACC GGC GTC CAC CGC GGC TCG AAG GTC CGG TTC ATC | 3853 |
| | Arg Ser Asn Tyr Thr Gly Val His Arg Gly Ser Lys Val Arg Phe Ile | |
| | 170 175 180 | |
| 5 | GAG TAC TTC AAC GGC ACG CGC CCG AGC CGG ATC CTC GTC GAC GTG CTC | 3901 |
| | Glu Tyr Phe Asn Gly Thr Arg Pro Ser Arg Ile Leu Val Asp Val Leu | |
| | 185 190 195 | |
| | CAG TCC GAG CAG GGC GCG TGG TTC CTG CGC AAG CGC AAC CGG AAC ATG | 3949 |
| 10 | Gln Ser Glu Gln Gly Ala Trp Phe Leu Arg Lys Arg Asn Arg Asn Met | |
| | 200 205 210 | |
| | GTC GTC GAG GTG TTC GAC GAC CTG CCC GAG CAC CCG AAC TTC CGG TGG | 3997 |
| | Val Val Glu Val Phe Asp Asp Leu Pro Glu His Pro Asn Phe Arg Trp | |
| 15 | 215 220 225 230 | |
| | CTG ACC GTC GCG CAG CTG CGG GCG ATG CTG CAC CAC GAC AAC GTG GTG | 4045 |
| | Leu Thr Val Ala Gln Leu Arg Ala Met Leu His His Asp Asn Val Val | |
| | 235 240 245 | |
| 20 | AAC ATG GAC CTG CGC ACC GTG CTG GCC TGC GTC CCG ACC GCC GTG GAG | 4093 |
| | Asn Met Asp Leu Arg Thr Val Leu Ala Cys Val Pro Thr Ala Val Glu | |
| | 250 255 260 | |
| 25 | CGG GAC CGG GCC GAC GAC GTG CTC GCG CGC CTG CCC GAG GGC TCG TTC | 4141 |
| | Arg Asp Arg Ala Asp Asp Val Leu Ala Arg Leu Pro Glu Gly Ser Phe | |
| | 265 270 275 | |
| | CAG GCC CGG CTG CTG CAC TCG TTC ATC GGC GCG GGC ACC CCG GCC AAC | 4189 |
| 30 | Gln Ala Arg Leu Leu His Ser Phe Ile Gly Ala Gly Thr Pro Ala Asn | |
| | 280 285 290 | |
| | AAC ATG AAC AGC CTG CTG AGC TGG ATC TCC GAC GTG CGC GCC AGG CGC | 4237 |
| | Asn Met Asn Ser Leu Leu Ser Trp Ile Ser Asp Val Arg Ala Arg Arg | |
| 35 | 295 300 305 310 | |
| | GAG TTC GTG CAG CGC GGC CGC CCG CTG CCC GAC ATC GAG CGC AGC GGG | 4285 |
| | Glu Phe Val Gln Arg Gly Arg Pro Leu Pro Asp Ile Glu Arg Ser Gly | |
| | 315 320 325 | |
| 40 | TGG ATC CGC CGC GAC GAC GGC ATC GAG CAC GAG GAG AAG AAG TAC TTC | 4333 |
| | Trp Ile Arg Arg Asp Asp Gly Ile Glu His Glu Glu Lys Lys Tyr Phe | |
| | 330 335 340 | |

45

| | | |
|----|---|------|
| | GAC GTC TTC GGC GTC ACG GTG GCG ACC AGC GAC CGC GAG GTC AAC TCG | 4381 |
| | Asp Val Phe Gly Val Thr Val Ala Thr Ser Asp Arg Glu Val Asn Ser | |
| | 345 350 355 | |
| 5 | TGG ATG CAG CCG CTG CTC TCG CCC GCC AAC AAC GGC CTG CTC GCC CTG | 4429 |
| | Trp Met Gln Pro Leu Leu Ser Pro Ala Asn Asn Gly Leu Leu Ala Leu | |
| | 360 365 370 | |
| | CTG GTC AAG GAC ATC GGC GGC ACG TTG CAC GCG CTC GTG CAG CTG CGC | 4477 |
| 10 | Leu Val Lys Asp Ile Gly Gly Thr Leu His Ala Leu Val Gln Leu Arg | |
| | 375 380 385 390 | |
| | ACC GAG GCG GGC GGG ATG GAC GTC GCC GAG CTG GCG CCT ACG GTG CAC | 4525 |
| 15 | Thr Glu Ala Gly Gly Met Asp Val Ala Glu Leu Ala Pro Thr Val His | |
| | 395 400 405 | |
| | TGC CAG CCC GAC AAC TAC GCC GAC GCG CCC GAG GAG TTC CGA CCG GCC | 4573 |
| | Cys Gln Pro Asp Asn Tyr Ala Asp Ala Pro Glu Glu Phe Arg Pro Ala | |
| | 410 415 420 | |
| 20 | TAT GTG GAC TAC GTG TTG AAC GTG CCG CGC TCG CAG GTC CGC TAC GAC | 4621 |
| | Tyr Val Asp Tyr Val Leu Asn Val Pro Arg Ser Gln Val Arg Tyr Asp | |
| | 425 430 435 | |
| 25 | GCA TGG CAC TCC GAG GAG GGC GGC CGG TTC TAC CGC AAC GAG AAC CGG | 4669 |
| | Ala Trp His Ser Glu Glu Gly Gly Arg Phe Tyr Arg Asn Glu Asn Arg | |
| | 440 445 450 | |
| | TAC ATG CTG ATC GAG GTG CCC GCC GAC TTC GAC GCC AGT GCC GCT CCC | 4717 |
| 30 | Tyr Met Leu Ile Glu Val Pro Ala Asp Phe Asp Ala Ser Ala Ala Pro | |
| | 455 460 465 470 | |
| | GAC CAC CGG TGG ATG ACC TTC GAC CAG ATC ACC TAC CTG CTC GGG CAC | 4765 |
| 35 | Asp His Arg Trp Met Thr Phe Asp Gln Ile Thr Tyr Leu Leu Gly His | |
| | 475 480 485 | |
| | AGC CAC TAC GTC AAC ATC CAG CTG CGC AGC ATC ATC GCG TGC GCC TCG | 4813 |
| | Ser His Tyr Val Asn Ile Gln Leu Arg Ser Ile Ile Ala Cys Ala Ser | |
| | 490 495 500 | |
| 40 | GCC GTC TAC ACC AGG ACC GCC GGA TGAAACGCGC GCTGACCGAC CTGGCGATCT | 4867 |
| | Ala Val Tyr Thr Arg Thr Ala Gly | |
| | 505 510 | |
| 45 | TCGGCGGCCC CGAGGCATTC CTGCACACCC TCTACGTGGG CAGGCCGACC GTCGGGGACC | 4927 |

| | | |
|----|--|------|
| | GGGAGCGGTT CTTCGCCCCG CTGGAGTGGG CGCTGAACAA CAACTGGCTG ACCAACGGCG | 4987 |
| | GACCACTGGT GCGCGAGTTC GAGGGCCGGG TCGCCGACCT GCGGGGTGTC CGCCACTGCG | 5047 |
| 5 | TGGCCACCTG CAACGCGACG GTCGCGCTGC AACTGGTGCT GCGCGCGAGC GACGTGTCCG | 5107 |
| | GCGAGGTCGT CATGCCTTCG ATGACGTTTCG CGGCCACCGC GCACGCGGCG AGCTGGCTGG | 5167 |
| | GGCTGGAACC GGTGTTCTGC GACGTGGACC CCGAGACCGG CCTGCTCGAC CCCGAGCACG | 5227 |
| 10 | TCGCGTCGCT GGTGACACCG CGGACGGGCG CGATCATCGG CGTGACCTG TGGGGCAGGC | 5287 |
| | CCGCTCCGGT CGAGGCGCTG GAGAAGATCG CCGCCGAGCA CCAGGTCAAA CTCTTCTTCG | 5347 |
| 15 | ACGCCGCGCA CGCGCTGGGC TGCACCGCCG GCGGGCGGCC GGTGCGCGCC TTCGGCAACG | 5407 |
| | CCGAGGTGTT CAGCTTCCAC GCCACGAAGG CGGTCACCTC GTTCGAGGGC GGCGCCATCG | 5467 |
| | TCACCGACGA CGGGCTGCTG GCCGACCGCA TCCGCGCCAT GCACAACTTC GGGATCGCAC | 5527 |
| 20 | CGGACAAGCT GGTGACCGAT GTCGGCACCA ACGGCAAGAT GAGCGAGTGC GCCGCGGCGA | 5587 |
| | TGGGCCTCAC CTCGCTCGAC GCCTTCGCCG AGACCAGGGT GCACAACCGC CTCAACCACG | 5647 |
| 25 | CGCTCTACTC CGACGAGCTC CGCGACGTGC GCGGCATATC CGTGACGCG TTCGATCCTG | 5707 |
| | GCGAGCAGAA CAACTACCAG TACGTGATCA TCTCGGTGGA CTCCGCGGCC ACCGGCATCG | 5767 |
| | ACCGCGACCA GTTGCAAGCG ATCCTGCGAG CGGAGAAGGT TGTGGCACAA CCCTACTTCT | 5827 |
| 30 | CCCCCGGGTG CCACCAGATG CAGCCGTACC GGACCGAGCC GCCGCTGCGG CTGGAGAACA | 5887 |
| | CCGAACAGCT CTCCGACCGG GTGCTCGCGC TGCCACCGG CCCC GCGGTG TCCAGCGAGG | 5947 |
| 35 | ACATCCGGCG GGTGTGCGAC ATCATCCGGC TCGCCGCCAC CAGCGGCGAG CTGATCAACG | 6007 |
| | CGCAATGGGA CCAGAGGACG CGCAACGGTT CGTGACGACC TCGGCCACAA GTGCCAGGAG | 6067 |
| | GTTCGCTCCC CG ATG AAC ACA ACT CGT ACG GCA ACC GCC CAG GAA GCG | 6115 |
| 40 | Met Asn Thr Thr Arg Thr Ala Thr Ala Gln Glu Ala | |
| | 1 5 10 | |
| | GGG GTC GCC GAC GCG GCG CGC CCG GAC GTC GAC CGG CGG GCG GTC GTG | 6163 |
| | Gly Val Ala Asp Ala Ala Arg Pro Asp Val Asp Arg Arg Ala Val Val | |
| 45 | 15 20 25 | |

23

| | | |
|----|---|------|
| | CGG GCG CTG AGC TCG GAG GTC TCC CGC GTC ACC GGC GCC GGT GAC GGT | 6211 |
| | Arg Ala Leu Ser Ser Glu Val Ser Arg Val Thr Gly Ala Gly Asp Gly | |
| | 30 35 40 | |
| 5 | GAC GCC GAC GTG CAG GCC GCC CGG CTC GCC GAC CTC GCC GCG CAC TAC | 6259 |
| | Asp Ala Asp Val Gln Ala Ala Arg Leu Ala Asp Leu Ala Ala His Tyr | |
| | 45 50 55 60 | |
| | GGG GCG CAC CCG TTC ACG CCG CTG GAG CAG ACG CGT GCG CCG CTC GGC | 6307 |
| 10 | Gly Ala His Pro Phe Thr Pro Leu Glu Gln Thr Arg Ala Arg Leu Gly | |
| | 65 70 75 | |
| | CTG GAC CGC GCG GAG TTC GCC CAC CTG CTC GAC CTG TTC GGC CGC ATC | 6355 |
| | Leu Asp Arg Ala Glu Phe Ala His Leu Leu Asp Leu Phe Gly Arg Ile | |
| 15 | 80 85 90 | |
| | CCG GAC CTG GGC ACC GCG GTG GAG CAC GGT CCG GCG GGC AAG TAC TGG | 6403 |
| | Pro Asp Leu Gly Thr Ala Val Glu His Gly Pro Ala Gly Lys Tyr Trp | |
| | 95 100 105 | |
| 20 | TCC AAC ACG ATC AAG CCG CTG GAC GCC GCA GGC GCA CTG GAC GCG GCG | 6451 |
| | Ser Asn Thr Ile Lys Pro Leu Asp Ala Ala Gly Ala Leu Asp Ala Ala | |
| | 110 115 120 | |
| 25 | GTC TAC CGC AAG CCT GCC TTC CCC TAC AGC GTC GGC CTG TAC CCC GGC | 6499 |
| | Val Tyr Arg Lys Pro Ala Phe Pro Tyr Ser Val Gly Leu Tyr Pro Gly | |
| | 125 130 135 140 | |
| | CCG ACG TGC ATG TTC CGC TGC CAC TTC TGC GTG CCG GTG ACC GGT GCC | 6547 |
| 30 | Pro Thr Cys Met Phe Arg Cys His Phe Cys Val Arg Val Thr Gly Ala | |
| | 145 150 155 | |
| | CGC TAC GAG GCC GCA TCG GTC CCG GCG GGC AAC GAG ACG CTG GCC GCG | 6595 |
| | Arg Tyr Glu Ala Ala Ser Val Pro Ala Gly Asn Glu Thr Leu Ala Ala | |
| 35 | 160 165 170 | |
| | ATC ATC GAC GAG GTG CCC ACG GAC AAC CCG AAG GCG ATG TAC ATG TCG | 6643 |
| | Ile Ile Asp Glu Val Pro Thr Asp Asn Pro Lys Ala Met Tyr Met Ser | |
| | 175 180 185 | |
| 40 | GGC GGC CTC GAG CCG CTG ACC AAC CCC GGT CTC GGC GAG CTG GTG TCG | 6691 |
| | Gly Gly Leu Glu Pro Leu Thr Asn Pro Gly Leu Gly Glu Leu Val Ser | |
| | 190 195 200 | |
| 45 | | |

24

| | | |
|----|---|------|
| | CAC GCC GCC GGG CGC GGT TTC GAC CTC ACC GTC TAC ACC AAC GCC TTC | 6739 |
| | His Ala Ala Gly Arg Gly Phe Asp Leu Thr Val Tyr Thr Asn Ala Phe | |
| | 205 210 215 220 | |
| 5 | GCC CTC ACC GAG CAG ACG CTG AAC CGC CAG CCC GGC CTG TGG GAG CTG | 6787 |
| | Ala Leu Thr Glu Gln Thr Leu Asn Arg Gln Pro Gly Leu Trp Glu Leu | |
| | 225 230 235 | |
| | GGC GCG ATC CGC ACG TCC CTC TAC GGG CTG AAC AAC GAC GAG TAC GAG | 6835 |
| 10 | Gly Ala Ile Arg Thr Ser Leu Tyr Gly Leu Asn Asn Asp Glu Tyr Glu | |
| | 240 245 250 | |
| | ACG ACC ACC GGC AAG CGC GGC GCT TTC GAA CGC GTC AAG AAG AAC CTG | 6883 |
| 15 | Thr Thr Thr Gly Lys Arg Gly Ala Phe Glu Arg Val Lys Lys Asn Leu | |
| | 255 260 265 | |
| | CAG GGC TTC CTG CGG ATG CGC GCC GAG CGG GAC GCG CCG ATC CGG CTC | 6931 |
| | Gln Gly Phe Leu Arg Met Arg Ala Glu Arg Asp Ala Pro Ile Arg Leu | |
| | 270 275 280 | |
| 20 | GGC TTC AAC CAC ATC ATC CTG CCG GGA CGG GCC GAC CGG CTC ACC GAC | 6979 |
| | Gly Phe Asn His Ile Ile Leu Pro Gly Arg Ala Asp Arg Leu Thr Asp | |
| | 285 290 295 300 | |
| 25 | CTC GTC GAC TTC ATC GCC GAG CTC AAC GAG TCC AGC CCG CAA CGG CCG | 7027 |
| | Leu Val Asp Phe Ile Ala Glu Leu Asn Glu Ser Ser Pro Gln Arg Pro | |
| | 305 310 315 | |
| | CTG GAC TTC GTG ACG GTG CGC GAG GAC TAC AGC GGC CGC GAC GAC GGC | 7075 |
| 30 | Leu Asp Phe Val Thr Val Arg Glu Asp Tyr Ser Gly Arg Asp Asp Gly | |
| | 320 325 330 | |
| | CGG CTG TCG GAC TCC GAG CGC AAC GAG CTG CGC GAG GGC CTG GTG CGG | 7123 |
| 35 | Arg Leu Ser Asp Ser Glu Arg Asn Glu Leu Arg Glu Gly Leu Val Arg | |
| | 335 340 345 | |
| | TTC GTC GAC TAC GCC GCC GAG CGG ACC CCG GGC ATG CAC ATC GAC CTG | 7171 |
| | Phe Val Asp Tyr Ala Ala Glu Arg Thr Pro Gly Met His Ile Asp Leu | |
| | 350 355 360 | |
| 40 | GGC TAC GCC CTG GAG AGC CTG CGG CGG GGT GTG GAC GCC GAG CTG CTG | 7219 |
| | Gly Tyr Ala Leu Glu Ser Leu Arg Arg Gly Val Asp Ala Glu Leu Leu | |
| | 365 370 375 380 | |

45

[illegible]

26

| | | |
|----|---|------|
| | ACC CCG CCG GGG CAG TGC AAG TAC GTC TAC TGC GCG CGC GGC CGG GCG | 7799 |
| | Thr Pro Pro Gly Gln Cys Lys Tyr Val Tyr Cys Ala Arg Gly Arg Ala | |
| | 60 65 70 | |
| 5 | CTC GAC GTC ATC GTC GAC ATC CGG GTC GGC TCG CCG ACG TTC GGG AAG | 7847 |
| | Leu Asp Val Ile Val Asp Ile Arg Val Gly Ser Pro Thr Phe Gly Lys | |
| | 75 80 85 90 | |
| | TGG GAC GCG GTG GAG ATG GAC ACC GAG CAC TTC CCG GCG GTC TAC TTC | 7895 |
| 10 | Trp Asp Ala Val Glu Met Asp Thr Glu His Phe Arg Ala Val Tyr Phe | |
| | 95 100 105 | |
| | CCC AGG GGC ACC GCG CAC GCC TTC CTC GCG CTT GAG GAC GAC ACC CTG | 7943 |
| | Pro Arg Gly Thr Ala His Ala Phe Leu Ala Leu Glu Asp Asp Thr Leu | |
| 15 | 110 115 120 | |
| | ATG TCG TAC CTG GTC AGC ACG CCG TAC GTG GCC GAG TAC GAG CAG GCG | 7991 |
| | Met Ser Tyr Leu Val Ser Thr Pro Tyr Val Ala Glu Tyr Glu Gln Ala | |
| | 125 130 135 | |
| 20 | | |
| | ATC GAC CCG TTC GAC CCC GCG CTG GGT CTG CCG TGG CCC GCG GAC CTG | 8039 |
| | Ile Asp Pro Phe Asp Pro Ala Leu Gly Leu Pro Trp Pro Ala Asp Leu | |
| | 140 145 150 | |
| 25 | GAG GTC GTG CTC TCC GAC CGC GAC ACG GTG GCC GTG GAC CTG GAG ACC | 8087 |
| | Glu Val Val Leu Ser Asp Arg Asp Thr Val Ala Val Asp Leu Glu Thr | |
| | 155 160 165 170 | |
| | GCC AGG CGG CGA GGG ATG CTG CCC GAC TAC GCC GAC TGC CTC GGC GAG | 8135 |
| 30 | Ala Arg Arg Arg Gly Met Leu Pro Asp Tyr Ala Asp Cys Leu Gly Glu | |
| | 175 180 185 | |
| | GAG CCC GCC AGC ACC GGC AGG TGAC | 8160 |
| | Glu Pro Ala Ser Thr Gly Arg | |
| 35 | 190 | |

(2) INFORMATIONS POUR LA SEQ ID NO: 7:

- 40 (i) CARACTERISTIQUES DE LA SEQUENCE:
- (A) LONGUEUR: 322 acides aminés
 - (B) TYPE: acide aminé
 - (D) CONFIGURATION: linéaire
- 45 (ii) TYPE DE MOLECULE: protéine
- (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 7:

27

Met Asn Gly Ile Ser Asp Ser Pro Arg Gln Leu Ile Thr Leu Leu Gly
 1 5 10 15
 Ala Ser Gly Phe Val Gly Ser Ala Val Leu Arg Glu Leu Arg Asp His
 5 20 25 30
 Pro Val Arg Leu Arg Ala Val Ser Arg Gly Gly Ala Pro Ala Val Pro
 35 40 45
 10 Pro Gly Ala Ala Glu Val Glu Asp Leu Arg Ala Asp Leu Leu Glu Pro
 50 55 60
 Gly Arg Ala Ala Ala Ala Ile Glu Asp Ala Asp Val Ile Val His Leu
 65 70 75 80
 15 Val Ala His Ala Ala Gly Gly Ser Thr Trp Arg Ser Ala Thr Ser Asp
 85 90 95
 Pro Glu Ala Glu Arg Val Asn Val Gly Leu Met His Asp Leu Val Gly
 20 100 105 110
 Ala Leu His Asp Arg Arg Arg Ser Thr Pro Pro Val Leu Leu Tyr Ala
 115 120 125
 25 Ser Thr Ala Gln Ala Ala Asn Pro Ser Ala Ala Ser Arg Tyr Ala Gln
 130 135 140
 Gln Lys Thr Glu Ala Glu Arg Ile Leu Arg Lys Ala Thr Asp Glu Gly
 145 150 155 160
 30 Arg Val Arg Gly Val Ile Leu Arg Leu Pro Ala Val Tyr Gly Gln Ser
 165 170 175
 Gly Pro Ser Gly Pro Met Gly Arg Gly Val Val Ala Ala Met Ile Arg
 35 180 185 190
 Arg Ala Leu Ala Gly Glu Pro Leu Thr Met Trp His Asp Gly Gly Val
 195 200 205
 40 Arg Arg Asp Leu Leu His Val Glu Asp Val Ala Thr Ala Phe Ala Ala
 210 215 220
 Ala Leu Glu His His Asp Ala Leu Ala Gly Gly Thr Trp Ala Leu Gly
 225 230 235 240
 45

28

Ala Asp Arg Ser Glu Pro Leu Gly Asp Ile Phe Arg Ala Val Ser Gly
 245 250 255

Ser Val Ala Arg Gln Thr Gly Ser Pro Ala Val Asp Val Val Thr Val
 5 260 265 270

Pro Ala Pro Glu His Ala Glu Ala Asn Asp Phe Arg Ser Asp Asp Ile
 275 280 285

10 Asp Ser Thr Glu Phe Arg Ser Arg Thr Gly Trp Arg Pro Arg Val Ser
 290 295 300

Leu Thr Asp Gly Ile Asp Arg Thr Val Ala Ala Leu Thr Pro Thr Glu
 305 310 315 320

15 Glu His

(2) INFORMATIONS POUR LA SEQ ID NO: 8:

20

(i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 415 acides aminés

(B) TYPE: acide aminé

(D) CONFIGURATION: linéaire

25

(ii) TYPE DE MOLECULE: protéine

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 8:

Met Arg Val Leu Leu Thr Ser Phe Ala His Arg Thr His Phe Gln Gly
 30 1 5 10 15

Leu Val Pro Leu Ala Trp Ala Leu Arg Thr Ala Gly His Asp Val Arg
 20 25 30

35 Val Ala Ala Gln Pro Ala Leu Thr Asp Ala Val Ile Gly Ala Gly Leu
 35 40 45

Thr Ala Val Pro Val Gly Ser Asp His Arg Leu Phe Asp Ile Val Pro
 50 55 60

40 Glu Val Ala Ala Gln Val His Arg Tyr Ser Phe Tyr Leu Asp Phe Tyr
 65 70 75 80

His Arg Glu Gln Glu Leu His Ser Trp Glu Phe Leu Leu Gly Met Gln
 45 85 90 95

29

Glu Ala Thr Ser Arg Trp Val Tyr Pro Val Val Asn Asn Asp Ser Phe
 100 105 110

Val Ala Glu Leu Val Asp Phe Ala Arg Asp Trp Arg Pro Asp Leu Val
 5 115 120 125

Leu Trp Glu Pro Phe Thr Phe Ala Gly Ala Val Ala Ala Arg Ala Cys
 130 135 140

10 Gly Ala Ala His Ala Arg Leu Leu Trp Gly Ser Asp Leu Thr Gly Tyr
 145 150 155 160

Phe Arg Gly Arg Phe Gln Ala Gln Arg Leu Arg Arg Pro Pro Glu Asp
 165 170 175

15 Arg Pro Asp Pro Leu Gly Thr Trp Leu Thr Glu Val Ala Gly Arg Phe
 180 185 190

Gly Val Glu Phe Gly Glu Asp Leu Ala Val Gly Gln Trp Ser Val Asp
 20 195 200 205

Gln Leu Pro Pro Ser Phe Arg Leu Asp Thr Gly Met Glu Thr Val Val
 210 215 220

25 Ala Arg Thr Leu Pro Tyr Asn Gly Ala Ser Val Val Pro Asp Trp Leu
 225 230 235 240

Lys Lys Gly Ser Ala Thr Arg Arg Ile Cys Ile Thr Gly Gly Phe Ser
 245 250 255

30 Gly Leu Gly Leu Ala Ala Asp Ala Asp Gln Phe Ala Arg Thr Leu Ala
 260 265 270

Gln Leu Ala Arg Phe Asp Gly Glu Ile Val Val Thr Gly Ser Gly Pro
 35 275 280 285

Asp Thr Ser Ala Val Pro Asp Asn Ile Arg Leu Val Asp Phe Val Pro
 290 295 300

40 Met Gly Val Leu Leu Gln Asn Cys Ala Ala Ile Ile His His Gly Gly
 305 310 315 320

Ala Gly Thr Trp Ala Thr Ala Leu His His Gly Ile Pro Gln Ile Ser
 325 330 335

45

30

Val Ala His Glu Trp Asp Cys Met Leu Arg Gly Gln Gln Thr Ala Glu
 340 345 350

Leu Gly Ala Gly Ile Tyr Leu Arg Pro Asp Glu Val Asp Ala Asp Ser
 5 355 360 365

Leu Ala Ser Ala Leu Thr Gln Val Val Glu Asp Pro Thr Tyr Thr Glu
 370 375 380

10 Asn Ala Val Lys Leu Arg Glu Glu Ala Leu Ser Asp Pro Thr Pro Gln
 385 390 395 400

Glu Ile Val Pro Arg Leu Glu Glu Leu Thr Arg Arg His Ala Gly
 405 410 415

15

(2) INFORMATIONS POUR LA SEQ ID NO: 9:

(i) CARACTERISTIQUES DE LA SEQUENCE:

20 (A) LONGUEUR: 237 acides aminés
 (B) TYPE: acide aminé
 (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: protéine

25 (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 9:

Met Tyr Glu Gly Gly Phe Ala Glu Leu Tyr Asp Arg Phe Tyr Arg Gly
 1 5 10 15

30 Arg Gly Lys Asp Tyr Ala Ala Glu Ala Ala Gln Val Ala Arg Leu Val
 20 25 30

Arg Asp Arg Leu Pro Ser Ala Ser Ser Leu Leu Asp Val Ala Cys Gly
 35 40 45

35 Thr Gly Thr His Leu Arg Arg Phe Ala Asp Leu Phe Asp Asp Val Thr
 50 55 60

Gly Leu Glu Leu Ser Ala Ala Met Ile Glu Val Ala Arg Pro Gln Leu
 40 65 70 75 80

Gly Gly Ile Pro Val Leu Gln Gly Asp Met Arg Asp Phe Ala Leu Asp
 85 90 95

45 Arg Glu Phe Asp Ala Val Thr Cys Met Phe Ser Ser Ile Gly His Met
 100 105 110

31

Arg Asp Gly Ala Glu Leu Asp Gln Ala Leu Ala Ser Phe Ala Arg His
 115 120 125
 Leu Ala Pro Gly Gly Val Val Val Val Glu Pro Trp Trp Phe Pro Glu
 5 130 135 140
 Asp Phe Leu Asp Gly Tyr Val Ala Gly Asp Val Val Arg Asp Gly Asp
 145 150 155 160
 10 Leu Thr Ile Ser Arg Val Ser His Ser Val Arg Ala Gly Gly Ala Thr
 165 170 175
 Arg Met Glu Ile His Trp Val Val Ala Asp Ala Val Asn Gly Pro Arg
 180 185 190
 15 His His Val Glu His Tyr Glu Ile Thr Leu Phe Glu Arg Gln Gln Tyr
 195 200 205
 Glu Lys Ala Phe Thr Ala Ala Gly Cys Ala Val Gln Tyr Leu Glu Gly
 20 210 215 220
 Gly Pro Ser Gly Arg Gly Leu Phe Val Gly Val Arg Gly
 225 230 235

25

(2) INFORMATIONS POUR LA SEQ ID NO: 10:

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 510 acides aminés
 (B) TYPE: acide aminé
 (D) CONFIGURATION: linéaire

30

(ii) TYPE DE MOLECULE: protéine

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 10:

35

Met Arg Val Leu Ile Asp Asn Ala Arg Arg Gln Gln Ala Glu Pro Ser
 1 5 10 15
 Thr Thr Pro Gln Gly Glu Ser Met Gly Asp Arg Thr Gly Asp Arg Thr
 40 20 25 30
 Ile Pro Glu Ser Ser Gln Thr Ala Thr Arg Phe Leu Leu Gly Asp Gly
 35 40 45
 45 Gly Ile Pro Thr Ala Thr Ala Glu Thr His Asp Trp Leu Thr Arg Asn
 50 55 60

32

Gly Ala Glu Gln Arg Leu Glu Val Ala Arg Val Pro Phe Ser Ala Met
 65 70 75 80

Asp Arg Trp Ser Phe Gln Pro Glu Asp Gly Arg Leu Ala His Glu Ser
 5 85 90 95

Gly Arg Phe Phe Ser Ile Glu Gly Leu His Val Arg Thr Asn Phe Gly
 100 105 110

10 Trp Arg Arg Asp Trp Ile Gln Pro Ile Ile Val Gln Pro Glu Ile Gly
 115 120 125

Phe Leu Gly Leu Ile Val Lys Glu Phe Asp Gly Val Leu His Val Leu
 130 135 140

15 Ala Gln Ala Lys Ala Glu Pro Gly Asn Ile Asn Ala Val Gln Leu Ser
 145 150 155 160

Pro Thr Leu Gln Ala Thr Arg Ser Asn Tyr Thr Gly Val His Arg Gly
 20 165 170 175

Ser Lys Val Arg Phe Ile Glu Tyr Phe Asn Gly Thr Arg Pro Ser Arg
 180 185 190

25 Ile Leu Val Asp Val Leu Gln Ser Glu Gln Gly Ala Trp Phe Leu Arg
 195 200 205

Lys Arg Asn Arg Asn Met Val Val Glu Val Phe Asp Asp Leu Pro Glu
 210 215 220

30 His Pro Asn Phe Arg Trp Leu Thr Val Ala Gln Leu Arg Ala Met Leu
 225 230 235 240

His His Asp Asn Val Val Asn Met Asp Leu Arg Thr Val Leu Ala Cys
 35 245 250 255

Val Pro Thr Ala Val Glu Arg Asp Arg Ala Asp Asp Val Leu Ala Arg
 260 265 270

40 Leu Pro Glu Gly Ser Phe Gln Ala Arg Leu Leu His Ser Phe Ile Gly
 275 280 285

Ala Gly Thr Pro Ala Asn Asn Met Asn Ser Leu Leu Ser Trp Ile Ser
 290 295 300

45

33

Asp Val Arg Ala Arg Arg Glu Phe Val Gln Arg Gly Arg Pro Leu Pro
 305 310 315 320

Asp Ile Glu Arg Ser Gly Trp Ile Arg Arg Asp Asp Gly Ile Glu His
 5 325 330 335

Glu Glu Lys Lys Tyr Phe Asp Val Phe Gly Val Thr Val Ala Thr Ser
 340 345 350

10 Asp Arg Glu Val Asn Ser Trp Met Gln Pro Leu Leu Ser Pro Ala Asn
 355 360 365

Asn Gly Leu Leu Ala Leu Leu Val Lys Asp Ile Gly Gly Thr Leu His
 370 375 380

15 Ala Leu Val Gln Leu Arg Thr Glu Ala Gly Gly Met Asp Val Ala Glu
 385 390 395 400

Leu Ala Pro Thr Val His Cys Gln Pro Asp Asn Tyr Ala Asp Ala Pro
 20 405 410 415

Glu Glu Phe Arg Pro Ala Tyr Val Asp Tyr Val Leu Asn Val Pro Arg
 420 425 430

25 Ser Gln Val Arg Tyr Asp Ala Trp His Ser Glu Glu Gly Gly Arg Phe
 435 440 445

Tyr Arg Asn Glu Asn Arg Tyr Met Leu Ile Glu Val Pro Ala Asp Phe
 450 455 460

30 Asp Ala Ser Ala Ala Pro Asp His Arg Trp Met Thr Phe Asp Gln Ile
 465 470 475 480

Thr Tyr Leu Leu Gly His Ser His Tyr Val Asn Ile Gln Leu Arg Ser
 35 485 490 495

Ile Ile Ala Cys Ala Ser Ala Val Tyr Thr Arg Thr Ala Gly
 500 505 510

40

(2) INFORMATIONS POUR LA SEQ ID NO: 11:

(i) CARACTERISTIQUES DE LA SEQUENCE:

- 45 (A) LONGUEUR: 489 acides aminés
 (B) TYPE: acide aminé
 (D) CONFIGURATION: linéaire

34

(ii) TYPE DE MOLECULE: protéine

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 11:

```

Met Asn Thr Thr Arg Thr Ala Thr Ala Gln Glu Ala Gly Val Ala Asp
5   1                               5           10           15

Ala Ala Arg Pro Asp Val Asp Arg Arg Ala Val Val Arg Ala Leu Ser
                20                25                30

10 Ser Glu Val Ser Arg Val Thr Gly Ala Gly Asp Gly Asp Ala Asp Val
                35                40                45

Gln Ala Ala Arg Leu Ala Asp Leu Ala Ala His Tyr Gly Ala His Pro
50                55                60
15

Phe Thr Pro Leu Glu Gln Thr Arg Ala Arg Leu Gly Leu Asp Arg Ala
65                70                75                80

Glu Phe Ala His Leu Leu Asp Leu Phe Gly Arg Ile Pro Asp Leu Gly
20                85                90                95

Thr Ala Val Glu His Gly Pro Ala Gly Lys Tyr Trp Ser Asn Thr Ile
                100                105                110

25 Lys Pro Leu Asp Ala Ala Gly Ala Leu Asp Ala Ala Val Tyr Arg Lys
                115                120                125

Pro Ala Phe Pro Tyr Ser Val Gly Leu Tyr Pro Gly Pro Thr Cys Met
130                135                140
30

Phe Arg Cys His Phe Cys Val Arg Val Thr Gly Ala Arg Tyr Glu Ala
145                150                155                160

Ala Ser Val Pro Ala Gly Asn Glu Thr Leu Ala Ala Ile Ile Asp Glu
35                165                170                175

Val Pro Thr Asp Asn Pro Lys Ala Met Tyr Met Ser Gly Gly Leu Glu
                180                185                190

40 Pro Leu Thr Asn Pro Gly Leu Gly Glu Leu Val Ser His Ala Ala Gly
                195                200                205

Arg Gly Phe Asp Leu Thr Val Tyr Thr Asn Ala Phe Ala Leu Thr Glu
210                215                220
45

```

35

Gln Thr Leu Asn Arg Gln Pro Gly Leu Trp Glu Leu Gly Ala Ile Arg
 225 230 235 240

5 Thr Ser Leu Tyr Gly Leu Asn Asn Asp Glu Tyr Glu Thr Thr Thr Gly
 245 250 255

Lys Arg Gly Ala Phe Glu Arg Val Lys Lys Asn Leu Gln Gly Phe Leu
 260 265 270

10 Arg Met Arg Ala Glu Arg Asp Ala Pro Ile Arg Leu Gly Phe Asn His
 275 280 285

Ile Ile Leu Pro Gly Arg Ala Asp Arg Leu Thr Asp Leu Val Asp Phe
 15 290 295 300

Ile Ala Glu Leu Asn Glu Ser Ser Pro Gln Arg Pro Leu Asp Phe Val
 305 310 315 320

20 Thr Val Arg Glu Asp Tyr Ser Gly Arg Asp Asp Gly Arg Leu Ser Asp
 325 330 335

Ser Glu Arg Asn Glu Leu Arg Glu Gly Leu Val Arg Phe Val Asp Tyr
 340 345 350

25 Ala Ala Glu Arg Thr Pro Gly Met His Ile Asp Leu Gly Tyr Ala Leu
 355 360 365

Glu Ser Leu Arg Arg Gly Val Asp Ala Glu Leu Leu Arg Ile Arg Pro
 30 370 375 380

Glu Thr Met Arg Pro Thr Ala His Pro Gln Val Ala Val Gln Ile Asp
 385 390 395 400

35 Leu Leu Gly Asp Val Tyr Leu Tyr Arg Glu Ala Gly Phe Pro Glu Leu
 405 410 415

Glu Gly Ala Thr Arg Tyr Ile Ala Gly Arg Val Thr Pro Ser Thr Ser
 420 425 430

40 Leu Arg Glu Val Val Glu Asn Phe Val Leu Glu Asn Glu Gly Val Gln
 435 440 445

Pro Arg Pro Gly Asp Glu Tyr Phe Leu Asp Gly Phe Asp Gln Ser Val
 45 450 455 460

36

Thr Ala Arg Leu Asn Gln Leu Glu Arg Asp Ile Ala Asp Gly Trp Glu
 465 470 475 480

Asp His Arg Gly Phe Leu Arg Gly Arg
 5 485

(2) INFORMATIONS POUR LA SEQ ID NO: 12:

10 (i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 193 acides aminés

(B) TYPE: acide aminé

(D) CONFIGURATION: linéaire

15 (ii) TYPE DE MOLECULE: protéine

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 12:

Met Ala Gly Gly Phe Glu Phe Thr Pro Asp Pro Lys Gln Asp Arg Arg
 1 5 10 15

20 Gly Leu Phe Val Ser Pro Leu Gln Asp Glu Ala Phe Val Gly Ala Val
 20 25 30

Gly His Arg Phe Pro Val Ala Gln Met Asn His Ile Val Ser Ala Arg
 25 35 40 45

Gly Val Leu Arg Gly Leu His Phe Thr Thr Thr Pro Pro Gly Gln Cys
 50 55 60

30 Lys Tyr Val Tyr Cys Ala Arg Gly Arg Ala Leu Asp Val Ile Val Asp
 65 70 75 80

Ile Arg Val Gly Ser Pro Thr Phe Gly Lys Trp Asp Ala Val Glu Met
 85 90 95

35 Asp Thr Glu His Phe Arg Ala Val Tyr Phe Pro Arg Gly Thr Ala His
 100 105 110

Ala Phe Leu Ala Leu Glu Asp Asp Thr Leu Met Ser Tyr Leu Val Ser
 40 115 120 125

Thr Pro Tyr Val Ala Glu Tyr Glu Gln Ala Ile Asp Pro Phe Asp Pro
 130 135 140

45 Ala Leu Gly Leu Pro Trp Pro Ala Asp Leu Glu Val Val Leu Ser Asp
 145 150 155 160

37

Arg Asp Thr Val Ala Val Asp Leu Glu Thr Ala Arg Arg Arg Gly Met
 165 170 175

Leu Pro Asp Tyr Ala Asp Cys Leu Gly Glu Glu Pro Ala Ser Thr Gly
 5 180 185 190

Arg

10 (2) INFORMATIONS POUR LA SEQ ID NO: 13:

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 1206 paires de bases
 (B) TYPE: nucléotide
 15 (C) NOMBRE DE BRINS: double
 (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN (génomique)

20 (vi) ORIGINE:

(A) ORGANISME: *Saccharopolyspora erythraea*

(ix) CARACTERISTIQUE:

- (A) NOM/CLE: CDS
 25 (B) EMBLEMMENT:1..1203
 (D) AUTRES INFORMATIONS:/function= "implique dans la
 biosynthese de la desosamine"
 /gene= "eryCIV"
 /note= "SEQ ID NO 6 DE 4837 A 6039"

30

(ix) CARACTERISTIQUE:

- (A) NOM/CLE: mat_peptide
 (B) EMBLEMMENT:1

35

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 13:

ATG AAA CGC GCG CTG ACC GAC CTG GCG ATC TTC GGC GGC CCC GAG GCA 48
 Met Lys Arg Ala Leu Thr Asp Leu Ala Ile Phe Gly Gly Pro Glu Ala
 40 1 5 10 15
 TTC CTG CAC ACC CTC TAC GTG GGC AGG CCG ACC GTC GGG GAC CGG GAG 96
 Phe Leu His Thr Leu Tyr Val Gly Arg Pro Thr Val Gly Asp Arg Glu
 20 25 30
 45

38

| | | |
|----|---|-----|
| | CGG TTC TTC GCC CGC CTG GAG TGG GCG CTG AAC AAC AAC TGG CTG ACC | 144 |
| | Arg Phe Phe Ala Arg Leu Glu Trp Ala Leu Asn Asn Asn Trp Leu Thr | |
| | 35 40 45 | |
| 5 | AAC GGC GGA CCA CTG GTG CGC GAG TTC GAG GGC CGG GTC GCC GAC CTG | 192 |
| | Asn Gly Gly Pro Leu Val Arg Glu Phe Glu Gly Arg Val Ala Asp Leu | |
| | 50 55 60 | |
| | GCG GGT GTC CGC CAC TGC GTG GCC ACC TGC AAC GCG ACG GTC GCG CTG | 240 |
| 10 | Ala Gly Val Arg His Cys Val Ala Thr Cys Asn Ala Thr Val Ala Leu | |
| | 65 70 75 80 | |
| | CAA CTG GTG CTG CGC GCG AGC GAC GTG TCC GGC GAG GTC GTC ATG CCT | 288 |
| | Gln Leu Val Leu Arg Ala Ser Asp Val Ser Gly Glu Val Val Met Pro | |
| 15 | 85 90 95 | |
| | TCG ATG ACG TTC GCG GCC ACC GCG CAC GCG GCG AGC TGG CTG GGG CTG | 336 |
| | Ser Met Thr Phe Ala Ala Thr Ala His Ala Ala Ser Trp Leu Gly Leu | |
| | 100 105 110 | |
| 20 | GAA CCG GTG TTC TGC GAC GTG GAC CCC GAG ACC GGC CTG CTC GAC CCC | 384 |
| | Glu Pro Val Phe Cys Asp Val Asp Pro Glu Thr Gly Leu Leu Asp Pro | |
| | 115 120 125 | |
| 25 | GAG CAC GTC GCG TCG CTG GTG ACA CCG CGG ACG GGC GCG ATC ATC GGC | 432 |
| | Glu His Val Ala Ser Leu Val Thr Pro Arg Thr Gly Ala Ile Ile Gly | |
| | 130 135 140 | |
| | GTG CAC CTG TGG GGC AGG CCC GCT CCG GTC GAG GCG CTG GAG AAG ATC | 480 |
| 30 | Val His Leu Trp Gly Arg Pro Ala Pro Val Glu Ala Leu Glu Lys Ile | |
| | 145 150 155 160 | |
| | GCC GCC GAG CAC CAG GTC AAA CTC TTC TTC GAC GCC GCG CAC GCG CTG | 528 |
| | Ala Ala Glu His Gln Val Lys Leu Phe Phe Asp Ala Ala His Ala Leu | |
| 35 | 165 170 175 | |
| | GGC TGC ACC GCC GGC GGG CGG CCG GTC GGC GCC TTC GGC AAC GCC GAG | 576 |
| | Gly Cys Thr Ala Gly Gly Arg Pro Val Gly Ala Phe Gly Asn Ala Glu | |
| | 180 185 190 | |
| 40 | GTG TTC AGC TTC CAC GCC ACG AAG GCG GTC ACC TCG TTC GAG GGC GGC | 624 |
| | Val Phe Ser Phe His Ala Thr Lys Ala Val Thr Ser Phe Glu Gly Gly | |
| | 195 200 205 | |

45

GCC ATC GTC ACC GAC GAC GGG CTG CTG GCC GAC CGC ATC CGC GCC ATG 672
 Ala Ile Val Thr Asp Asp Gly Leu Leu Ala Asp Arg Ile Arg Ala Met
 210 215 220

5 CAC AAC TTC GGG ATC GCA CCG GAC AAG CTG GTG ACC GAT GTC GGC ACC 720
 His Asn Phe Gly Ile Ala Pro Asp Lys Leu Val Thr Asp Val Gly Thr
 225 230 235 240

AAC GGC AAG ATG AGC GAG TGC GCC GCG GCG ATG GGC CTC ACC TCG CTC 768
 10 Asn Gly Lys Met Ser Glu Cys Ala Ala Ala Met Gly Leu Thr Ser Leu
 245 250 255

GAC GCC TTC GCC GAG ACC AGG GTG CAC AAC CGC CTC AAC CAC GCG CTC 816
 Asp Ala Phe Ala Glu Thr Arg Val His Asn Arg Leu Asn His Ala Leu
 15 260 265 270

TAC TCC GAC GAG CTC CGC GAC GTG CGC GGC ATA TCC GTG CAC GCG TTC 864
 Tyr Ser Asp Glu Leu Arg Asp Val Arg Gly Ile Ser Val His Ala Phe
 275 280 285

20 GAT CCT GGC GAG CAG AAC AAC TAC CAG TAC GTG ATC ATC TCG GTG GAC 912
 Asp Pro Gly Glu Gln Asn Asn Tyr Gln Tyr Val Ile Ile Ser Val Asp
 290 295 300

25 TCC GCG GCC ACC GGC ATC GAC CGC GAC CAG TTG CAG GCG ATC CTG CGA 960
 Ser Ala Ala Thr Gly Ile Asp Arg Asp Gln Leu Gln Ala Ile Leu Arg
 305 310 315 320

GCG GAG AAG GTT GTG GCA CAA CCC TAC TTC TCC CCC GGG TGC CAC CAG 1008
 30 Ala Glu Lys Val Val Ala Gln Pro Tyr Phe Ser Pro Gly Cys His Gln
 325 330 335

ATG CAG CCG TAC CGG ACC GAG CCG CCG CTG CGG CTG GAG AAC ACC GAA 1056
 Met Gln Pro Tyr Arg Thr Glu Pro Pro Leu Arg Leu Glu Asn Thr Glu
 35 340 345 350

CAG CTC TCC GAC CGG GTG CTC GCG CTG CCC ACC GGC CCC GCG GTG TCC 1104
 Gln Leu Ser Asp Arg Val Leu Ala Leu Pro Thr Gly Pro Ala Val Ser
 355 360 365

40 AGC GAG GAC ATC CGG CGG GTG TGC GAC ATC ATC CGG CTC GCC GCC ACC 1152
 Ser Glu Asp Ile Arg Arg Val Cys Asp Ile Ile Arg Leu Ala Ala Thr
 370 375 380

40

AGC GGC GAG CTG ATC AAC GCG CAA TGG GAC CAG AGG ACG CGC AAC GGT 1200
 Ser Gly Glu Leu Ile Asn Ala Gln Trp Asp Gln Arg Thr Arg Asn Gly
 385 390 395 400

5 TCG TGA 1206
 Ser

(2) INFORMATIONS POUR LA SEQ ID NO: 14:

10 (i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 401 acides aminés

(B) TYPE: acide aminé

(D) CONFIGURATION: linéaire

15 (ii) TYPE DE MOLECULE: protéine

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 14:

Met Lys Arg Ala Leu Thr Asp Leu Ala Ile Phe Gly Gly Pro Glu Ala
 1 5 10 15

20 Phe Leu His Thr Leu Tyr Val Gly Arg Pro Thr Val Gly Asp Arg Glu
 20 25 30

Arg Phe Phe Ala Arg Leu Glu Trp Ala Leu Asn Asn Asn Trp Leu Thr
 25 35 40 45

Asn Gly Gly Pro Leu Val Arg Glu Phe Glu Gly Arg Val Ala Asp Leu
 50 55 60

30 Ala Gly Val Arg His Cys Val Ala Thr Cys Asn Ala Thr Val Ala Leu
 65 70 75 80

Gln Leu Val Leu Arg Ala Ser Asp Val Ser Gly Glu Val Val Met Pro
 85 90 95

35 Ser Met Thr Phe Ala Ala Thr Ala His Ala Ala Ser Trp Leu Gly Leu
 100 105 110

Glu Pro Val Phe Cys Asp Val Asp Pro Glu Thr Gly Leu Leu Asp Pro
 40 115 120 125

Glu His Val Ala Ser Leu Val Thr Pro Arg Thr Gly Ala Ile Ile Gly
 130 135 140

45 Val His Leu Trp Gly Arg Pro Ala Pro Val Glu Ala Leu Glu Lys Ile
 145 150 155 160

41

Ala Ala Glu His Gln Val Lys Leu Phe Phe Asp Ala Ala His Ala Leu
165 170 175

Gly Cys Thr Ala Gly Gly Arg Pro Val Gly Ala Phe Gly Asn Ala Glu
5 180 185 190

Val Phe Ser Phe His Ala Thr Lys Ala Val Thr Ser Phe Glu Gly Gly
195 200 205

10 Ala Ile Val Thr Asp Asp Gly Leu Leu Ala Asp Arg Ile Arg Ala Met
210 215 220

His Asn Phe Gly Ile Ala Pro Asp Lys Leu Val Thr Asp Val Gly Thr
225 230 235 240

15 Asn Gly Lys Met Ser Glu Cys Ala Ala Ala Met Gly Leu Thr Ser Leu
245 250 255

Asp Ala Phe Ala Glu Thr Arg Val His Asn Arg Leu Asn His Ala Leu
20 260 265 270

Tyr Ser Asp Glu Leu Arg Asp Val Arg Gly Ile Ser Val His Ala Phe
275 280 285

25 Asp Pro Gly Glu Gln Asn Asn Tyr Gln Tyr Val Ile Ile Ser Val Asp
290 295 300

Ser Ala Ala Thr Gly Ile Asp Arg Asp Gln Leu Gln Ala Ile Leu Arg
305 310 315 320

30 Ala Glu Lys Val Val Ala Gln Pro Tyr Phe Ser Pro Gly Cys His Gln
325 330 335

Met Gln Pro Tyr Arg Thr Glu Pro Pro Leu Arg Leu Glu Asn Thr Glu
35 340 345 350

Gln Leu Ser Asp Arg Val Leu Ala Leu Pro Thr Gly Pro Ala Val Ser
355 360 365

40 Ser Glu Asp Ile Arg Arg Val Cys Asp Ile Ile Arg Leu Ala Ala Thr
370 375 380

Ser Gly Glu Leu Ile Asn Ala Gln Trp Asp Gln Arg Thr Arg Asn Gly
385 390 395 400

45 Ser

(2) INFORMATIONS POUR LA SEQ ID NO: 15:

- (i) CARACTERISTIQUES DE LA SEQUENCE:
- 5 (A) LONGUEUR: 6093 paires de bases
(B) TYPE: nucléotide
(C) NOMBRE DE BRINS: double
(D) CONFIGURATION: linéaire
- 10 (ii) TYPE DE MOLECULE: ADN (génomique)
- (vi) ORIGINE:
(A) ORGANISME: Streptomyces antibioticus
- 15 (ix) CARACTERISTIQUE:
(A) NOM/CLE: CDS
(B) EMBLACEMENT:184..1386
(D) AUTRES INFORMATIONS:/gene= "oleP1"
- 20 (ix) CARACTERISTIQUE:
(A) NOM/CLE: CDS
(B) EMBLACEMENT:1437..2714
(D) AUTRES INFORMATIONS:/function= "glycosylation de
8,8a-desoxyoleandolide"
25 /gene= "oleG1"
/transl_except= (pos: 1437 .. 1439, aa: Met)
- (ix) CARACTERISTIQUE:
(A) NOM/CLE: CDS
30 (B) EMBLACEMENT:2722..3999
(D) AUTRES INFORMATIONS:/function= "glycosylation de
8,8a-desoxyoleandolide"
/gene= "oleG2"
- 35 (ix) CARACTERISTIQUE:
(A) NOM/CLE: CDS
(B) EMBLACEMENT:4810..5967
(D) AUTRES INFORMATIONS:/gene= "oleY"
- 40 (ix) CARACTERISTIQUE:
(A) NOM/CLE: mat_peptide
(B) EMBLACEMENT:184
- 45 (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 15:

| | | |
|----|---|-----|
| | GCATGCCCCG CTTTCCTCCC CCTCTCCGAA CGCATCGACG ACCCGATCCC CCTCAGGGAC | 60 |
| | CGGTGAAGGA GCGTGTGCA CTCATGCAGG ACATGCAAGG CGTACAGCCC GAACCAGCCA | 120 |
| 5 | GTGTGGAACA CGCGGCGGAC GCAGCTCGAA CAGAGCGAAC GGCGCACGGA AGCCGCCCAG | 180 |
| | GAG ATG GAG GAC AGC GAA CTG GGG CGC CGC CTG CAG ATG CTC CGC GGC | 228 |
| | Met Glu Asp Ser Glu Leu Gly Arg Arg Leu Gln Met Leu Arg Gly | |
| | 1 5 10 15 | |
| 10 | ATG CAG TGG GTC TTC GGC GCC AAC GGC GAT CCG TAC GCC CGG CTG CTG | 276 |
| | Met Gln Trp Val Phe Gly Ala Asn Gly Asp Pro Tyr Ala Arg Leu Leu | |
| | 20 25 30 | |
| 15 | TGT GGC ATG GAG GAT GAC CCG TCA CCT TTC TAC GAC GCG ATA CGG ACC | 324 |
| | Cys Gly Met Glu Asp Asp Pro Ser Pro Phe Tyr Asp Ala Ile Arg Thr | |
| | 35 40 45 | |
| | CTG GGC GAG CTG CAC CGG AGC AGG ACC GGA GCC TGG GTC ACC GCC GAC | 372 |
| 20 | Leu Gly Glu Leu His Arg Ser Arg Thr Gly Ala Trp Val Thr Ala Asp | |
| | 50 55 60 | |
| | CCC GGG CTC GGG GGC CGC ATC CTC GCC GAC CGG AAG GCT CGG TGC CCG | 420 |
| | Pro Gly Leu Gly Gly Arg Ile Leu Ala Asp Arg Lys Ala Arg Cys Pro | |
| 25 | 65 70 75 | |
| | GAA GGC TCG TGG CCG GTG CGG GCG AAG ACC GAC GGG CTG GAG CAG TAC | 468 |
| | Glu Gly Ser Trp Pro Val Arg Ala Lys Thr Asp Gly Leu Glu Gln Tyr | |
| | 80 85 90 95 | |
| 30 | GTG CTG CCC GGG CAC CAG GCG TTC CTG CGG CTG GAG CGC GAG GAG GCC | 516 |
| | Val Leu Pro Gly His Gln Ala Phe Leu Arg Leu Glu Arg Glu Glu Ala | |
| | 100 105 110 | |
| 35 | GAG CGA CTG CGG GAG GTC GCG GCG CCG GTG CTG GGG GCC GCG GCG GTC | 564 |
| | Glu Arg Leu Arg Glu Val Ala Ala Pro Val Leu Gly Ala Ala Ala Val | |
| | 115 120 125 | |
| | GAC GCG TGG CGC CCG CTG ATC GAC GAG GTC TGC GCG GGG CTC GCG AAG | 612 |
| 40 | Asp Ala Trp Arg Pro Leu Ile Asp Glu Val Cys Ala Gly Leu Ala Lys | |
| | 130 135 140 | |
| | GGG CTG CCG GAC ACG TTC GAC CTG GTC GAG GAG TAC GCG GGG CTG GTG | 660 |
| | Gly Leu Pro Asp Thr Phe Asp Leu Val Glu Glu Tyr Ala Gly Leu Val | |
| 45 | 145 150 155 | |

44

| | | |
|----|---|------|
| | CCG GTC GAG GTG CTG GCG CGG ATC TGG GGC GTC CCG GAG GAG GAC CGC | 708 |
| | Pro Val Glu Val Leu Ala Arg Ile Trp Gly Val Pro Glu Glu Asp Arg | |
| | 160 165 170 175 | |
| 5 | GCC CGG TTC GGG CGT GAC TGC CGG GCG CTC GCT CCC GCG CTG GAC AGC | 756 |
| | Ala Arg Phe Gly Arg Asp Cys Arg Ala Leu Ala Pro Ala Leu Asp Ser | |
| | 180 185 190 | |
| | CTC CTG TGT CCC CAG CAG TTG GCG CTG AGC AAG GAC ATG GCG TCC GCC | 804 |
| 10 | Leu Leu Cys Pro Gln Gln Leu Ala Leu Ser Lys Asp Met Ala Ser Ala | |
| | 195 200 205 | |
| | CTG GAG GAC CTG CGT CTC CTC TTC GAC GGC CTC GAC GCG ACG CCG CGC | 852 |
| | Leu Glu Asp Leu Arg Leu Leu Phe Asp Gly Leu Asp Ala Thr Pro Arg | |
| 15 | 210 215 220 | |
| | CTC GCC GGC CCC GCC GAC GGT GAC GGA ACG GCC GTG GCC ATG CTC ACC | 900 |
| | Leu Ala Gly Pro Ala Asp Gly Asp Gly Thr Ala Val Ala Met Leu Thr | |
| | 225 230 235 | |
| 20 | GTT CTG CTC TGC ACG GAG CCG GTG ACC ACG GCG ATC GGG AAC ACC GTG | 948 |
| | Val Leu Leu Cys Thr Glu Pro Val Thr Thr Ala Ile Gly Asn Thr Val | |
| | 240 245 250 255 | |
| 25 | CTC GGG CTC CTT CCC GGG CAG TGG CCC GTG CCC TGC ACC GGC CGG GTG | 996 |
| | Leu Gly Leu Leu Pro Gly Gln Trp Pro Val Pro Cys Thr Gly Arg Val | |
| | 260 265 270 | |
| | GCT GCC GGG CAG GTT GCC GGG CAG GCG CTG CAC CGG GCG GTG TCG TAC | 1044 |
| 30 | Ala Ala Gly Gln Val Ala Gly Gln Ala Leu His Arg Ala Val Ser Tyr | |
| | 275 280 285 | |
| | CGT ATC GCG ACG CGG TTC GCC CGG GAG GAC CTG GAG TTG GCG GGC TGC | 1092 |
| | Arg Ile Ala Thr Arg Phe Ala Arg Glu Asp Leu Glu Leu Ala Gly Cys | |
| 35 | 290 295 300 | |
| | GAG GTC AAG TCC GGT GAC GAG GTG GTG GTC CTG GCC GGA GCG ATC GGC | 1140 |
| | Glu Val Lys Ser Gly Asp Glu Val Val Val Leu Ala Gly Ala Ile Gly | |
| | 305 310 315 | |
| 40 | CGG AAC GGA CCG TCC GCA GCC GCC CCG CCT GCC CCA CCG GGC CCA GCG | 1188 |
| | Arg Asn Gly Pro Ser Ala Ala Ala Pro Pro Ala Pro Pro Gly Pro Ala | |
| | 320 325 330 335 | |

45

| | | | | | | | | | | | | | | | | | |
|----|-----|-----|------------|------------|------------|-----------|-------------|------|-----|-----|-----|-----|-----|-----|-----|-----|------|
| | GCC | CCG | CCC | GCC | CCG | TCG | GTC | TTC | GGT | GCC | GCC | GCC | TTC | GAG | AAC | GCG | 1236 |
| | Ala | Pro | Pro | Ala | Pro | Ser | Val | Phe | Gly | Ala | Ala | Ala | Phe | Glu | Asn | Ala | |
| | | | | | 340 | | | | | 345 | | | | | 350 | | |
| 5 | CTG | GCC | GAA | CCC | CTC | GTC | CGG | GCT | GTG | ACG | GGA | GCG | GCC | CTC | CAG | GCC | 1284 |
| | Leu | Ala | Glu | Pro | Leu | Val | Arg | Ala | Val | Thr | Gly | Ala | Ala | Leu | Gln | Ala | |
| | | | | | 355 | | | | | 360 | | | | | 365 | | |
| | CTC | GCG | GAG | GGG | CCC | CCC | CGG | CTG | ACG | GCG | GCG | GGA | CCC | GTC | GTA | CGA | 1332 |
| 10 | Leu | Ala | Glu | Gly | Pro | Pro | Arg | Leu | Thr | Ala | Ala | Gly | Pro | Val | Val | Arg | |
| | | | | | 370 | | | | | 375 | | | | | 380 | | |
| | CGG | CGG | CGT | TCC | CCT | GTC | GTC | GGC | GGG | CTG | CAC | CGG | GCT | CCG | GTG | GCC | 1380 |
| | Arg | Arg | Arg | Ser | Pro | Val | Val | Gly | Gly | Leu | His | Arg | Ala | Pro | Val | Ala | |
| 15 | | | 385 | | | | | 390 | | | | | | 395 | | | |
| | GCC | GCA | TGAGCATCGC | GTCGAACGGC | GCGCGCTCGG | CCCCCGCCG | GCCCCCTGCGC | 1436 | | | | | | | | | |
| | Ala | Ala | | | | | | 400 | | | | | | | | | |
| 20 | GTG | ATG | ATG | ACC | ACC | TTC | GCG | GCC | AAC | ACG | CAC | TTC | CAG | CCG | CTG | GTT | 1484 |
| | Met | Met | Met | Thr | Thr | Phe | Ala | Ala | Asn | Thr | His | Phe | Gln | Pro | Leu | Val | |
| | 1 | | | | 5 | | | | | 10 | | | | | 15 | | |
| 25 | CCC | CTG | GCC | TGG | GCA | CTG | CGG | ACA | GCC | GGG | CAC | GAG | GTG | CGC | GTG | GTG | 1532 |
| | Pro | Leu | Ala | Trp | Ala | Leu | Arg | Thr | Ala | Gly | His | Glu | Val | Arg | Val | Val | |
| | | | | | 20 | | | | | 25 | | | | | 30 | | |
| | AGC | CAG | CCC | TCG | CTG | AGC | GAC | GTG | GTG | ACG | CAG | GCG | GGG | CTC | ACC | TCG | 1580 |
| 30 | Ser | Gln | Pro | Ser | Leu | Ser | Asp | Val | Val | Thr | Gln | Ala | Gly | Leu | Thr | Ser | |
| | | | | | 35 | | | | | 40 | | | | | 45 | | |
| | GTC | CCG | GTG | GGC | ACC | GAG | GCT | CCG | GTC | GAG | CAG | TTC | GCG | GCG | ACC | TGG | 1628 |
| | Val | Pro | Val | Gly | Thr | Glu | Ala | Pro | Val | Glu | Gln | Phe | Ala | Ala | Thr | Trp | |
| 35 | | | 50 | | | | | 55 | | | | | | 60 | | | |
| | GGC | GAC | GAT | GCC | TAC | ATC | GGC | GTC | AAC | AGC | ATC | GAC | TTC | ACC | GGC | AAC | 1676 |
| | Gly | Asp | Asp | Ala | Tyr | Ile | Gly | Val | Asn | Ser | Ile | Asp | Phe | Thr | Gly | Asn | |
| | | | 65 | | | | | 70 | | | | | | 75 | | 80 | |
| 40 | GAC | CCC | GGC | CTG | TGG | ACG | TGG | CCG | TAC | CTC | CTG | GGC | ATG | GAG | ACC | ATG | 1724 |
| | Asp | Pro | Gly | Leu | Trp | Thr | Trp | Pro | Tyr | Leu | Leu | Gly | Met | Glu | Thr | Met | |
| | | | | | | | | 85 | | | | | | 90 | | 95 | |

BNSDOCID: <WO_9905283A2.1 >

46

| | | |
|----|---|------|
| | CTG GTG CCG GCC TTC TAC GAG TTG CTG AAC AAC GAG TCC TTC GTG GAC | 1772 |
| | Leu Val Pro Ala Phe Tyr Glu Leu Leu Asn Asn Glu Ser Phe Val Asp | |
| | 100 105 110 | |
| 5 | GGC GTA GTC GAG TTC GCC CGT GAC TGG CGG CCC GAC CTG GTG ATC TGG | 1820 |
| | Gly Val Val Glu Phe Ala Arg Asp Trp Arg Pro Asp Leu Val Ile Trp | |
| | 115 120 125 | |
| | GAG CCG CTG ACG TTC GCC GGC GCG GTG GCG GCG CGC GTC ACC GGC GCG | 1868 |
| 10 | Glu Pro Leu Thr Phe Ala Gly Ala Val Ala Ala Arg Val Thr Gly Ala | |
| | 130 135 140 | |
| | GCC CAC GCC CGG CTG CCG TGG GGG CAG GAG ATC ACC CTG CGC GGG CGG | 1916 |
| | Ala His Ala Arg Leu Pro Trp Gly Gln Glu Ile Thr Leu Arg Gly Arg | |
| 15 | 145 150 155 160 | |
| | CAG GCG TTC CTC GCC GAG CGT GCC CTG CAA CCG TTC GAG CAC CGG GAG | 1964 |
| | Gln Ala Phe Leu Ala Glu Arg Ala Leu Gln Pro Phe Glu His Arg Glu | |
| | 165 170 175 | |
| 20 | GAT CCC ACG GCC GAG TGG CTG GGC CGC ATG CTC GAC CGG TAC GGC TGC | 2012 |
| | Asp Pro Thr Ala Glu Trp Leu Gly Arg Met Leu Asp Arg Tyr Gly Cys | |
| | 180 185 190 | |
| 25 | TCG TTC GAC GAG GAG ATG GTC ACC GGG CAG TGG ACC ATC GAC ACG CTG | 2060 |
| | Ser Phe Asp Glu Glu Met Val Thr Gly Gln Trp Thr Ile Asp Thr Leu | |
| | 195 200 205 | |
| | CCG CGC AGC ATG CGG CTG GAG CTG TCC GAG GAG CTG CGC ACC CTG GAC | 2108 |
| 30 | Pro Arg Ser Met Arg Leu Glu Leu Ser Glu Glu Leu Arg Thr Leu Asp | |
| | 210 215 220 | |
| | ATG CGG TAC GTG CCG TAC AAC GGA CCG GCG GTC GTA CCC CCC TGG GTG | 2156 |
| | Met Arg Tyr Val Pro Tyr Asn Gly Pro Ala Val Val Pro Pro Trp Val | |
| 35 | 225 230 235 240 | |
| | TGG GAA CCG TGC GAG CGG CCC CGG GTC TGT CTG ACG ATC GGC ACC TCC | 2204 |
| | Trp Glu Pro Cys Glu Arg Pro Arg Val Cys Leu Thr Ile Gly Thr Ser | |
| | 245 250 255 | |
| 40 | CAG CGT GAC TCC GGC CGG GAC CAT GTC CCC CTC GAC CAC CTG CTC GAC | 2252 |
| | Gln Arg Asp Ser Gly Arg Asp His Val Pro Leu Asp His Leu Leu Asp | |
| | 260 265 270 | |

45

47

| | | |
|----|---|------|
| | TCC CTC GCC GAC GTG GAC GCG GAG ATC GTG GCC ACG CTC GAC ACC ACC | 2300 |
| | Ser Leu Ala Asp Val Asp Ala Glu Ile Val Ala Thr Leu Asp Thr Thr | |
| | 275 280 285 | |
| 5 | CAG CAG GAG CGC CTG CGG GGC GCG GCC CCC GGC AAC GTC CGG CTG GTG | 2348 |
| | Gln Gln Glu Arg Leu Arg Gly Ala Ala Pro Gly Asn Val Arg Leu Val | |
| | 290 295 300 | |
| | GAC TTC GTC CCG CTG CAC GCG CTG ATG CCG ACC TGC TCG GCG ATC GTG | 2396 |
| 10 | Asp Phe Val Pro Leu His Ala Leu Met Pro Thr Cys Ser Ala Ile Val | |
| | 305 310 315 320 | |
| | CAC CAC GGT GGT CCG GGC ACG TGG TCG ACG GCG GCG CTC CAC GGC GTC | 2444 |
| | His His Gly Gly Pro Gly Thr Trp Ser Thr Ala Ala Leu His Gly Val | |
| 15 | 325 330 335 | |
| | CCG CAG ATC ATC CTG GAC ACC TCG TGG GAC ACA CCG GTG CGG GCG CAG | 2492 |
| | Pro Gln Ile Ile Leu Asp Thr Ser Trp Asp Thr Pro Val Arg Ala Gln | |
| | 340 345 350 | |
| 20 | CGC ATG CAG CAA CTC GGG GCG GGC CTG TCG ATG CCG GTG GGG GAA CTG | 2540 |
| | Arg Met Gln Gln Leu Gly Ala Gly Leu Ser Met Pro Val Gly Glu Leu | |
| | 355 360 365 | |
| 25 | GGC GTC GAG GCG CTG CGG GAC CGG GTC CTG CGG CTG CTG GGG GAG CCG | 2588 |
| | Gly Val Glu Ala Leu Arg Asp Arg Val Leu Arg Leu Leu Gly Glu Pro | |
| | 370 375 380 | |
| | GAG TTC CGC GCG GGC GCC GAG CGG ATC CGG GCC GAG ATG CTC GCG ATG | 2636 |
| 30 | Glu Phe Arg Ala Gly Ala Glu Arg Ile Arg Ala Glu Met Leu Ala Met | |
| | 385 390 395 400 | |
| | CCC GCC CCC GGT GAC GTC GTA CCG GAC CTG GAA CGA CTC ACC GCG GAG | 2684 |
| | Pro Ala Pro Gly Asp Val Val Pro Asp Leu Glu Arg Leu Thr Ala Glu | |
| 35 | 405 410 415 | |
| | CAT GCC ACC GGC GCG ATG GCG GGA AGG CGG TGAGACG ATG CGC GTA CTG | 2733 |
| | His Ala Thr Gly Ala Met Ala Gly Arg Arg Met Arg Val Leu | |
| | 420 425 1 | |
| 40 | CTG ACC TGC TTC GCC AAC GAC ACC CAC TTC CAC GGG CTG GTG CCG CTG | 2781 |
| | Leu Thr Cys Phe Ala Asn Asp Thr His Phe His Gly Leu Val Pro Leu | |
| | 5 10 15 20 | |

45

48

GCG TGG GCG CTG CGG GCC GCC GGG CAC GAA GTC CGC GTG GCC AGT CAG 2829
 Ala Trp Ala Leu Arg Ala Ala Gly His Glu Val Arg Val Ala Ser Gln
 25 30 35

5 CCC GCC CTG TCC GAC ACG ATC ACC CAA GCG GGA CTG ACC GCG GTG CCC 2877
 Pro Ala Leu Ser Asp Thr Ile Thr Gln Ala Gly Leu Thr Ala Val Pro
 40 45 50

GTG GGC CGG GAC ACC GCC TTC CTG GAG CTG ATG GGG GAG ATC GGC GCG 2925
 10 Val Gly Arg Asp Thr Ala Phe Leu Glu Leu Met Gly Glu Ile Gly Ala
 55 60 65

GAC GTC CAG AAG TAC TCC ACC GGC ATC GAC CTG GGC GTC CGC GCG GAG 2973
 Asp Val Gln Lys Tyr Ser Thr Gly Ile Asp Leu Gly Val Arg Ala Glu
 15 70 75 80

CTG ACG AGC TGG GAG TAC CTG CTC GGC ATG CAC ACG ACC CTG GTG CCC 3021
 Leu Thr Ser Trp Glu Tyr Leu Leu Gly Met His Thr Thr Leu Val Pro
 85 90 95 100

20 ACG TTC TAC TCG CTG GTC AAC GAC GAG CCG TTC GTC GAC GGG CTC GTC 3069
 Thr Phe Tyr Ser Leu Val Asn Asp Glu Pro Phe Val Asp Gly Leu Val
 105 110 115

25 GCG CTG ACC CGG GCC TGG CGG CCC GAC CTC ATC CTG TGG GAG CAC TTC 3117
 Ala Leu Thr Arg Ala Trp Arg Pro Asp Leu Ile Leu Trp Glu His Phe
 120 125 130

AGC TTC GCC GGG GCG TTG GCG GCG CGG GCC ACC GGC ACG CCC CAC GCC 3165
 30 Ser Phe Ala Gly Ala Leu Ala Ala Arg Ala Thr Gly Thr Pro His Ala
 135 140 145

CGC GTG CTG TGG GGG TCG GAC CTC ATC GTC CGG TTC CGC CGG GAC TTC 3213
 Arg Val Leu Trp Gly Ser Asp Leu Ile Val Arg Phe Arg Arg Asp Phe
 35 150 155 160

CTC GCG GAG CGG GCG AAC CGG CCC GCC GAG CAC CGC GAG GAC CCC ATG 3261
 Leu Ala Glu Arg Ala Asn Arg Pro Ala Glu His Arg Glu Asp Pro Met
 165 170 175 180

40 GCG GAG TGG CTG GGC TGG GCG GCC GAA CGG CTG GGC TCC ACC TTC GAC 3309
 Ala Glu Trp Leu Gly Trp Ala Ala Glu Arg Leu Gly Ser Thr Phe Asp
 185 190 195

45

| | | |
|----|---|------|
| | GAG GAG CTG GTG ACC GGG CAG TGG ACG ATC GAC CCG CTG CCG CGG AGC | 3357 |
| | Glu Glu Leu Val Thr Gly Gln Trp Thr Ile Asp Pro Leu Pro Arg Ser | |
| | 200 205 210 | |
| 5 | ATG CGG CTG CCC ACC GGG ACG ACG ACG GTG CCG ATG CGG TAC GTG CCG | 3405 |
| | Met Arg Leu Pro Thr Gly Thr Thr Thr Val Pro Met Arg Tyr Val Pro | |
| | 215 220 225 | |
| | TAC AAC GGG CGG GCC GTG GTC CCC GCA TGG GTC CGG CAG CGT GCG CGG | 3453 |
| 10 | Tyr Asn Gly Arg Ala Val Val Pro Ala Trp Val Arg Gln Arg Ala Arg | |
| | 230 235 240 | |
| | CGG CCC CGG ATC TGC CTG ACG CTC GGT GTG TCG GCC CGG CAG ACC CTG | 3501 |
| | Arg Pro Arg Ile Cys Leu Thr Leu Gly Val Ser Ala Arg Gln Thr Leu | |
| 15 | 245 250 255 260 | |
| | GGC GAC GGC GTG TCG CTG GCG GAG GTG CTG GCC GCG CTG GGC GAC GTG | 3549 |
| | Gly Asp Gly Val Ser Leu Ala Glu Val Leu Ala Ala Leu Gly Asp Val | |
| | 265 270 275 | |
| 20 | GAC GCG GAG ATC GTG GCC ACG CTG GAC GCC TCC CAG CGC AAG CTC CTG | 3597 |
| | Asp Ala Glu Ile Val Ala Thr Leu Asp Ala Ser Gln Arg Lys Leu Leu | |
| | 280 285 290 | |
| 25 | GGG CCG GTG CCG GAC AAC GTC CGG CTG GTG GAC TTC GTG CCC CTG CAC | 3645 |
| | Gly Pro Val Pro Asp Asn Val Arg Leu Val Asp Phe Val Pro Leu His | |
| | 295 300 305 | |
| | GCC CTG ATG CCG ACC TGT TCG GCG ATC GTG CAC CAC GGC GGC GCC GGT | 3693 |
| 30 | Ala Leu Met Pro Thr Cys Ser Ala Ile Val His His Gly Gly Ala Gly | |
| | 310 315 320 | |
| | ACC TGG CTG ACG GCC GCC GTC CAC GGC GTC CCG CAG ATC GTC CTC GGT | 3741 |
| | Thr Trp Leu Thr Ala Ala Val His Gly Val Pro Gln Ile Val Leu Gly | |
| 35 | 325 330 335 340 | |
| | GAC CTC TGG GAC AAC CTG CTG CGC GCC CGG CAG ACA CAG GCC GCG GGC | 3789 |
| | Asp Leu Trp Asp Asn Leu Leu Arg Ala Arg Gln Thr Gln Ala Ala Gly | |
| | 345 350 355 | |
| 40 | GCG GGC CTG TTC ATC CAT CCG TCC GAG GTC ACC GCG GCC GGG CTC GGT | 3837 |
| | Ala Gly Leu Phe Ile His Pro Ser Glu Val Thr Ala Ala Gly Leu Gly | |
| | 360 365 370 | |

50

GAG GGC GTG CGC CGG GTG CTG ACG GAC CCT TCC ATC CGG GCC GCC GCA 3885
 Glu Gly Val Arg Arg Val Leu Thr Asp Pro Ser Ile Arg Ala Ala Ala
 375 380 385

5 CAG CGC GTC CGG GAC GAG ATG AAT GCA GAG CCG ACG CCG GGC GAG GTC 3933
 Gln Arg Val Arg Asp Glu Met Asn Ala Glu Pro Thr Pro Gly Glu Val
 390 395 400

GTC ACG GTG CTG GAG CGG CTC GCC GCG AGC GGC GGA CGC GGA CGA GGA 3981
 10 Val Thr Val Leu Glu Arg Leu Ala Ala Ser Gly Gly Arg Gly Arg Gly
 405 410 415 420

GGC GGG AAC CAT GCG GGC TGACACGGAG CCGACCACCG GGTACGAGGA 4029
 Gly Gly Asn His Ala Gly
 15 425

CGAGTTCGCC GAGATCTACG ACGCCGTGTA CCGGGGCCGG GGCAAGGACT ACGCCGGCGA 4089

GGCGAAGGAC GTGGCGGACC TCGTGCGCGA CCGGGTGCCG GACGCGTCCT CCCTCCTGGA 4149
 20 CGTGGCCTGC GGCACGGGCG CGCACCTGCG GCACTTCGCC ACGCTCTTCG ACGACGCCCCG 4209

CCGTCTCGAA CTGTCCGCGA GCATGCTGGA CATCGCCCGC TCCCGCATGC CGGGCGTGCC 4269

25 GCTGCACCAA GGGGACATGC GATCCTTCGA CCTGGGGCCA CGCGTCTCCG CGGTACCTG 4329

CATGTTTACG TCCGTGGGCC ACCTGGCCAC CACCGCCGAA CTCGACGCGA CGCTGCGGTG 4389

CTTGCCCCGG CACACCCGGC CCGGCGGCGT GGCCGTCATC GAACCGTGGT GGTTCGCCGA 4449
 30 GACCTTCACC GACGGCTACG TGGCGGGTGA CATCGTACGC GTCGACGGCC GGACCATCTC 4509

CCGGGTGTCC CACTCGGTAC GGGACGGCGG CGCCACCCGC ATGGAGATCC ACTACGTGAT 4569

35 CGCCGACGCC GAGCACGGTC CCCGGCACCT GGTCGAGCAC CACCGCATCA CGCTGTTCCC 4629

GCGGCATGCG TACACGGCCG CGTACGAGAA GCGGGGCTAC ACCGTCGAGT ACCTCGACGG 4689

CGGGCCCTCG GGCCGGGGGC TGTTCGTCGG CACCCGGACG TGAACCCGCC CGCGCACCGC 4749
 40 CCGATCACCC TGCTCAACGC CGTTCACACG GATCACCGGA CCACGCGAAG GACCTTTCAC 4809

ATG TCG TAC GAC GAC CAC GCG GTG CTG GAA GCG ATA CTG CGG TGC GCC 4857
 Met Ser Tyr Asp Asp His Ala Val Leu Glu Ala Ile Leu Arg Cys Ala
 45 1 5 10 15

51

| | | |
|----|---|------|
| | GGA GGT GAC GAG CGC TTC CTG CTG AAC ACC GTC GAG GAA TGG GGA GCC | 4905 |
| | Gly Gly Asp Glu Arg Phe Leu Leu Asn Thr Val Glu Glu Trp Gly Ala | |
| | 20 25 30 | |
| 5 | GCC GAG ATC ACC GCG GCG CTC GTG GAC GAG TTG CTG TTC CGC TGC GAG | 4953 |
| | Ala Glu Ile Thr Ala Ala Leu Val Asp Glu Leu Leu Phe Arg Cys Glu | |
| | 35 40 45 | |
| | ATC CCG CAG GTG GGC GGT GAG GCG TTC ATC GGC CTG GAC GTC CTG CAC | 5001 |
| 10 | Ile Pro Gln Val Gly Gly Glu Ala Phe Ile Gly Leu Asp Val Leu His | |
| | 50 55 60 | |
| | GGC GCC GAC CGG ATC AGC CAT GTG CTG CAG GTG ACG GAC GGC AAG CCG | 5049 |
| | Gly Ala Asp Arg Ile Ser His Val Leu Gln Val Thr Asp Gly Lys Pro | |
| 15 | 65 70 75 80 | |
| | GTC ACG TCG GCG GAA CCG GCC GGC CAG GAA CTG GGC GGC CGT ACC TGG | 5097 |
| | Val Thr Ser Ala Glu Pro Ala Gly Gln Glu Leu Gly Gly Arg Thr Trp | |
| | 85 90 95 | |
| 20 | AGT TCA CGC TCA GCG ACC CTC CTG CGG GAG CTG TTC GGC CCG CCG TCC | 5145 |
| | Ser Ser Arg Ser Ala Thr Leu Leu Arg Glu Leu Phe Gly Pro Pro Ser | |
| | 100 105 110 | |
| 25 | GGC CGC ACC GCG GGG GGC TTC GGC GTC TCC TTC CTG CCC GAC CTG CGC | 5193 |
| | Gly Arg Thr Ala Gly Gly Phe Gly Val Ser Phe Leu Pro Asp Leu Arg | |
| | 115 120 125 | |
| | GGC CCG CGG ACC ATG GAG GGC GCG GCC CTG GCC GCC CGC GCC ACC AAC | 5241 |
| 30 | Gly Pro Arg Thr Met Glu Gly Ala Ala Leu Ala Ala Arg Ala Thr Asn | |
| | 130 135 140 | |
| | GTG GTG CTG CAC GCG ACG ACC AAC GAG ACG CCC CCA CTG GAC CGG CTG | 5289 |
| | Val Val Leu His Ala Thr Thr Asn Glu Thr Pro Pro Leu Asp Arg Leu | |
| 35 | 145 150 155 160 | |
| | GCC CTG CGC TAC GAG TCC GAC AAG TGG GGC GGC GTC CAC TGG TTC ACC | 5337 |
| | Ala Leu Arg Tyr Glu Ser Asp Lys Trp Gly Gly Val His Trp Phe Thr | |
| | 165 170 175 | |
| 40 | GGC CAC TAC GAC CGG CAC CTG CGG GCC GTG CGC GAC CAG GCG GTG CGG | 5385 |
| | Gly His Tyr Asp Arg His Leu Arg Ala Val Arg Asp Gln Ala Val Arg | |
| | 180 185 190 | |

45

52

| | | |
|----|---|------|
| | ATC CTG GAG ATC GGC ATC GGC GGC TAC GAC GAC CTG CTG CCG AGC GGC | 5433 |
| | Ile Leu Glu Ile Gly Ile Gly Gly Tyr Asp Asp Leu Leu Pro Ser Gly | |
| | 195 200 205 | |
| 5 | GCC TCA CTG AAG ATG TGG AAG CGC TAC TTC CCG CGC GGC CTG GTC TTC | 5481 |
| | Ala Ser Leu Lys Met Trp Lys Arg Tyr Phe Pro Arg Gly Leu Val Phe | |
| | 210 215 220 | |
| | GGC GTG GAC ATC TTC GAC AGT CGG CGT GCG ACC AGC CGC GTG TCA AGA | 5529 |
| 10 | Gly Val Asp Ile Phe Asp Ser Arg Arg Ala Thr Ser Arg Val Ser Arg | |
| | 225 230 235 240 | |
| | CGC TCC GCG GCC CGG CAG GAC GAC CCG GAG TTC ATG CGC CGC GTC GCC | 5577 |
| | Arg Ser Ala Ala Arg Gln Asp Asp Pro Glu Phe Met Arg Arg Val Ala | |
| 15 | 245 250 255 | |
| | GAG GAG CAC GGG CCG TTC GAC GTC ATC ATC GAC GAC GGC AGC CAC ATC | 5625 |
| | Glu Glu His Gly Pro Phe Asp Val Ile Ile Asp Asp Gly Ser His Ile | |
| | 260 265 270 | |
| 20 | | |
| | AAC GCA CAC ATG CGG ACG TCG TTC TCG GTG ATG TTC CCC CAC CTG CGC | 5673 |
| | Asn Ala His Met Arg Thr Ser Phe Ser Val Met Phe Pro His Leu Arg | |
| | 275 280 285 | |
| 25 | AAC GGC GGC TTC TAC GTC ATC GAG GAC ACC TTC ACC TCC TAC TGG CCC | 5721 |
| | Asn Gly Gly Phe Tyr Val Ile Glu Asp Thr Phe Thr Ser Tyr Trp Pro | |
| | 290 295 300 | |
| | GGG TAC GGA GGG CCA TCC GGA GCC CGG TGC CCG TCC GGA ACA ACC GCG | 5769 |
| 30 | Gly Tyr Gly Gly Pro Ser Gly Ala Arg Cys Pro Ser Gly Thr Thr Ala | |
| | 305 310 315 320 | |
| | CTG GAG ATG GTC AAG GGA CTG ATC GAC TCG GTG CAC TAC GAG GAG CGG | 5817 |
| | Leu Glu Met Val Lys Gly Leu Ile Asp Ser Val His Tyr Glu Glu Arg | |
| 35 | 325 330 335 | |
| | CCG GAC GGC GCG GCC ACG GCC GAC TAC ATC GCC AGG AAC CTC GTC GGG | 5865 |
| | Pro Asp Gly Ala Ala Thr Ala Asp Tyr Ile Ala Arg Asn Leu Val Gly | |
| | 340 345 350 | |
| 40 | | |
| | CTG CAC GCC TAC CAA ACG ACC TCG TCT TCC TCG AGA AGG GCG ATC AAC | 5913 |
| | Leu His Ala Tyr Gln Thr Thr Ser Ser Ser Ser Arg Arg Ala Ile Asn | |
| | 355 360 365 | |

45

53

AAG GAG GGC GGC ATC CCC CAC ACC GTG CCC CGG GAG CCG TTC TGG AAC 5961
 Lys Glu Gly Gly Ile Pro His Thr Val Pro Arg Glu Pro Phe Trp Asn
 370 375 380

5 GAC AAC TAGCCACGGC CGCAACCAGA GCCGGAACC GCACCACTGT CCGCGCCACC 6017
 Asp Asn
 385

TCGGAACCAC CTCCAGCAAA GGACACACCG CTGTGACCGA TACGCACACC GGACCGACAC 6077
 10
 CGGCCGACGC GGTACC 6093

(2) INFORMATIONS POUR LA SEQ ID NO: 16:

15

(i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 401 acides aminés

(B) TYPE: acide aminé

(D) CONFIGURATION: linéaire

20

(ii) TYPE DE MOLECULE: protéine

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 16:

Met Glu Asp Ser Glu Leu Gly Arg Arg Leu Gln Met Leu Arg Gly Met
 25 1 5 10 15

Gln Trp Val Phe Gly Ala Asn Gly Asp Pro Tyr Ala Arg Leu Leu Cys
 20 25 30

30 Gly Met Glu Asp Asp Pro Ser Pro Phe Tyr Asp Ala Ile Arg Thr Leu
 35 40 45

Gly Glu Leu His Arg Ser Arg Thr Gly Ala Trp Val Thr Ala Asp Pro
 50 55 60

35 Gly Leu Gly Gly Arg Ile Leu Ala Asp Arg Lys Ala Arg Cys Pro Glu
 65 70 75 80

Gly Ser Trp Pro Val Arg Ala Lys Thr Asp Gly Leu Glu Gln Tyr Val
 40 85 90 95

Leu Pro Gly His Gln Ala Phe Leu Arg Leu Glu Arg Glu Glu Ala Glu
 100 105 110

45 Arg Leu Arg Glu Val Ala Ala Pro Val Leu Gly Ala Ala Val Asp
 115 120 125

54

Ala Trp Arg Pro Leu Ile Asp Glu Val Cys Ala Gly Leu Ala Lys Gly
 130 135 140

Leu Pro Asp Thr Phe Asp Leu Val Glu Glu Tyr Ala Gly Leu Val Pro
 5 145 150 155 160

Val Glu Val Leu Ala Arg Ile Trp Gly Val Pro Glu Glu Asp Arg Ala
 165 170 175

10 Arg Phe Gly Arg Asp Cys Arg Ala Leu Ala Pro Ala Leu Asp Ser Leu
 180 185 190

Leu Cys Pro Gln Gln Leu Ala Leu Ser Lys Asp Met Ala Ser Ala Leu
 195 200 205

15 Glu Asp Leu Arg Leu Leu Phe Asp Gly Leu Asp Ala Thr Pro Arg Leu
 210 215 220

Ala Gly Pro Ala Asp Gly Asp Gly Thr Ala Val Ala Met Leu Thr Val
 20 225 230 235 240

Leu Leu Cys Thr Glu Pro Val Thr Thr Ala Ile Gly Asn Thr Val Leu
 245 250 255

25 Gly Leu Leu Pro Gly Gln Trp Pro Val Pro Cys Thr Gly Arg Val Ala
 260 265 270

Ala Gly Gln Val Ala Gly Gln Ala Leu His Arg Ala Val Ser Tyr Arg
 275 280 285

30 Ile Ala Thr Arg Phe Ala Arg Glu Asp Leu Glu Leu Ala Gly Cys Glu
 290 295 300

Val Lys Ser Gly Asp Glu Val Val Val Leu Ala Gly Ala Ile Gly Arg
 35 305 310 315 320

Asn Gly Pro Ser Ala Ala Ala Pro Pro Ala Pro Pro Gly Pro Ala Ala
 325 330 335

40 Pro Pro Ala Pro Ser Val Phe Gly Ala Ala Ala Phe Glu Asn Ala Leu
 340 345 350

Ala Glu Pro Leu Val Arg Ala Val Thr Gly Ala Ala Leu Gln Ala Leu
 355 360 365

45

55

Ala Glu Gly Pro Pro Arg Leu Thr Ala Ala Gly Pro Val Val Arg Arg
 370 375 380

Arg Arg Ser Pro Val Val Gly Gly Leu His Arg Ala Pro Val Ala Ala
 5 385 390 395 400

Ala

10 (2) INFORMATIONS POUR LA SEQ ID NO: 17:

(i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 426 acides aminés

(B) TYPE: acide aminé

15 (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: protéine

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 17:

20 Met Met Met Thr Thr Phe Ala Ala Asn Thr His Phe Gln Pro Leu Val
 1 5 10 15

Pro Leu Ala Trp Ala Leu Arg Thr Ala Gly His Glu Val Arg Val Val
 20 25 30

25 Ser Gln Pro Ser Leu Ser Asp Val Val Thr Gln Ala Gly Leu Thr Ser
 35 40 45

Val Pro Val Gly Thr Glu Ala Pro Val Glu Gln Phe Ala Ala Thr Trp
 30 50 55 60

Gly Asp Asp Ala Tyr Ile Gly Val Asn Ser Ile Asp Phe Thr Gly Asn
 65 70 75 80

35 Asp Pro Gly Leu Trp Thr Trp Pro Tyr Leu Leu Gly Met Glu Thr Met
 85 90 95

Leu Val Pro Ala Phe Tyr Glu Leu Leu Asn Asn Glu Ser Phe Val Asp
 100 105 110

40 Gly Val Val Glu Phe Ala Arg Asp Trp Arg Pro Asp Leu Val Ile Trp
 115 120 125

Glu Pro Leu Thr Phe Ala Gly Ala Val Ala Ala Arg Val Thr Gly Ala
 45 130 135 140

56

Ala His Ala Arg Leu Pro Trp Gly Gln Glu Ile Thr Leu Arg Gly Arg
 145 150 155 160
 Gln Ala Phe Leu Ala Glu Arg Ala Leu Gln Pro Phe Glu His Arg Glu
 5 165 170 175
 Asp Pro Thr Ala Glu Trp Leu Gly Arg Met Leu Asp Arg Tyr Gly Cys
 180 185 190
 10 Ser Phe Asp Glu Glu Met Val Thr Gly Gln Trp Thr Ile Asp Thr Leu
 195 200 205
 Pro Arg Ser Met Arg Leu Glu Leu Ser Glu Glu Leu Arg Thr Leu Asp
 210 215 220
 15 Met Arg Tyr Val Pro Tyr Asn Gly Pro Ala Val Val Pro Pro Trp Val
 225 230 235 240
 Trp Glu Pro Cys Glu Arg Pro Arg Val Cys Leu Thr Ile Gly Thr Ser
 20 245 250 255
 Gln Arg Asp Ser Gly Arg Asp His Val Pro Leu Asp His Leu Leu Asp
 260 265 270
 25 Ser Leu Ala Asp Val Asp Ala Glu Ile Val Ala Thr Leu Asp Thr Thr
 275 280 285
 Gln Gln Glu Arg Leu Arg Gly Ala Ala Pro Gly Asn Val Arg Leu Val
 290 295 300
 30 Asp Phe Val Pro Leu His Ala Leu Met Pro Thr Cys Ser Ala Ile Val
 305 310 315 320
 His His Gly Gly Pro Gly Thr Trp Ser Thr Ala Ala Leu His Gly Val
 35 325 330 335
 Pro Gln Ile Ile Leu Asp Thr Ser Trp Asp Thr Pro Val Arg Ala Gln
 340 345 350
 40 Arg Met Gln Gln Leu Gly Ala Gly Leu Ser Met Pro Val Gly Glu Leu
 355 360 365
 Gly Val Glu Ala Leu Arg Asp Arg Val Leu Arg Leu Leu Gly Glu Pro
 370 375 380
 45

57

Glu Phe Arg Ala Gly Ala Glu Arg Ile Arg Ala Glu Met Leu Ala Met
 385 390 395 400

Pro Ala Pro Gly Asp Val Val Pro Asp Leu Glu Arg Leu Thr Ala Glu
 5 405 410 415

His Ala Thr Gly Ala Met Ala Gly Arg Arg
 420 425

10

(2) INFORMATIONS POUR LA SEQ ID NO: 18:

(i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 426 acides aminés

(B) TYPE: acide aminé

(D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: protéine

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 18:

20

Met Arg Val Leu Leu Thr Cys Phe Ala Asn Asp Thr His Phe His Gly
 1 5 10 15

Leu Val Pro Leu Ala Trp Ala Leu Arg Ala Ala Gly His Glu Val Arg
 25 20 25 30

Val Ala Ser Gln Pro Ala Leu Ser Asp Thr Ile Thr Gln Ala Gly Leu
 35 40 45

30 Thr Ala Val Pro Val Gly Arg Asp Thr Ala Phe Leu Glu Leu Met Gly
 50 55 60

Glu Ile Gly Ala Asp Val Gln Lys Tyr Ser Thr Gly Ile Asp Leu Gly
 65 70 75 80

35

Val Arg Ala Glu Leu Thr Ser Trp Glu Tyr Leu Leu Gly Met His Thr
 85 90 95

Thr Leu Val Pro Thr Phe Tyr Ser Leu Val Asn Asp Glu Pro Phe Val
 40 100 105 110

Asp Gly Leu Val Ala Leu Thr Arg Ala Trp Arg Pro Asp Leu Ile Leu
 115 120 125

45 Trp Glu His Phe Ser Phe Ala Gly Ala Leu Ala Ala Arg Ala Thr Gly
 130 135 140

58

Thr Pro His Ala Arg Val Leu Trp Gly Ser Asp Leu Ile Val Arg Phe
 145 150 155 160

Arg Arg Asp Phe Leu Ala Glu Arg Ala Asn Arg Pro Ala Glu His Arg
 5 165 170 175

Glu Asp Pro Met Ala Glu Trp Leu Gly Trp Ala Ala Glu Arg Leu Gly
 180 185 190

10 Ser Thr Phe Asp Glu Glu Leu Val Thr Gly Gln Trp Thr Ile Asp Pro
 195 200 205

Leu Pro Arg Ser Met Arg Leu Pro Thr Gly Thr Thr Thr Val Pro Met
 210 215 220

15 Arg Tyr Val Pro Tyr Asn Gly Arg Ala Val Val Pro Ala Trp Val Arg
 225 230 235 240

Gln Arg Ala Arg Arg Pro Arg Ile Cys Leu Thr Leu Gly Val Ser Ala
 20 245 250 255

Arg Gln Thr Leu Gly Asp Gly Val Ser Leu Ala Glu Val Leu Ala Ala
 260 265 270

25 Leu Gly Asp Val Asp Ala Glu Ile Val Ala Thr Leu Asp Ala Ser Gln
 275 280 285

Arg Lys Leu Leu Gly Pro Val Pro Asp Asn Val Arg Leu Val Asp Phe
 290 295 300

30 Val Pro Leu His Ala Leu Met Pro Thr Cys Ser Ala Ile Val His His
 305 310 315 320

Gly Gly Ala Gly Thr Trp Leu Thr Ala Ala Val His Gly Val Pro Gln
 35 325 330 335

Ile Val Leu Gly Asp Leu Trp Asp Asn Leu Leu Arg Ala Arg Gln Thr
 340 345 350

40 Gln Ala Ala Gly Ala Gly Leu Phe Ile His Pro Ser Glu Val Thr Ala
 355 360 365

Ala Gly Leu Gly Glu Gly Val Arg Arg Val Leu Thr Asp Pro Ser Ile
 370 375 380

45

59

Arg Ala Ala Ala Gln Arg Val Arg Asp Glu Met Asn Ala Glu Pro Thr
 385 390 395 400

Pro Gly Glu Val Val Thr Val Leu Glu Arg Leu Ala Ala Ser Gly Gly
 5 405 410 415

Arg Gly Arg Gly Gly Gly Asn His Ala Gly
 420 425

10

(2) INFORMATIONS POUR LA SEQ ID NO: 19:

(i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 386 acides aminés
 (B) TYPE: acide aminé
 (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: protéine

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 19:

20

Met Ser Tyr Asp Asp His Ala Val Leu Glu Ala Ile Leu Arg Cys Ala
 1 5 10 15

Gly Gly Asp Glu Arg Phe Leu Leu Asn Thr Val Glu Glu Trp Gly Ala
 25 20 25 30

Ala Glu Ile Thr Ala Ala Leu Val Asp Glu Leu Leu Phe Arg Cys Glu
 35 40 45

30 Ile Pro Gln Val Gly Gly Glu Ala Phe Ile Gly Leu Asp Val Leu His
 50 55 60

Gly Ala Asp Arg Ile Ser His Val Leu Gln Val Thr Asp Gly Lys Pro
 65 70 75 80

35

Val Thr Ser Ala Glu Pro Ala Gly Gln Glu Leu Gly Gly Arg Thr Trp
 85 90 95

Ser Ser Arg Ser Ala Thr Leu Leu Arg Glu Leu Phe Gly Pro Pro Ser
 40 100 105 110

Gly Arg Thr Ala Gly Gly Phe Gly Val Ser Phe Leu Pro Asp Leu Arg
 115 120 125

45 Gly Pro Arg Thr Met Glu Gly Ala Ala Leu Ala Ala Arg Ala Thr Asn
 130 135 140

60

Val Val Leu His Ala Thr Thr Asn Glu Thr Pro Pro Leu Asp Arg Leu
 145 150 155 160

Ala Leu Arg Tyr Glu Ser Asp Lys Trp Gly Gly Val His Trp Phe Thr
 5 165 170 175

Gly His Tyr Asp Arg His Leu Arg Ala Val Arg Asp Gln Ala Val Arg
 180 185 190

10 Ile Leu Glu Ile Gly Ile Gly Gly Tyr Asp Asp Leu Leu Pro Ser Gly
 195 200 205

Ala Ser Leu Lys Met Trp Lys Arg Tyr Phe Pro Arg Gly Leu Val Phe
 210 215 220

15 Gly Val Asp Ile Phe Asp Ser Arg Arg Ala Thr Ser Arg Val Ser Arg
 225 230 235 240

Arg Ser Ala Ala Arg Gln Asp Asp Pro Glu Phe Met Arg Arg Val Ala
 20 245 250 255

Glu Glu His Gly Pro Phe Asp Val Ile Ile Asp Asp Gly Ser His Ile
 260 265 270

25 Asn Ala His Met Arg Thr Ser Phe Ser Val Met Phe Pro His Leu Arg
 275 280 285

Asn Gly Gly Phe Tyr Val Ile Glu Asp Thr Phe Thr Ser Tyr Trp Pro
 290 295 300

30 Gly Tyr Gly Gly Pro Ser Gly Ala Arg Cys Pro Ser Gly Thr Thr Ala
 305 310 315 320

Leu Glu Met Val Lys Gly Leu Ile Asp Ser Val His Tyr Glu Glu Arg
 35 325 330 335

Pro Asp Gly Ala Ala Thr Ala Asp Tyr Ile Ala Arg Asn Leu Val Gly
 340 345 350

40 Leu His Ala Tyr Gln Thr Thr Ser Ser Ser Ser Arg Arg Ala Ile Asn
 355 360 365

Lys Glu Gly Gly Ile Pro His Thr Val Pro Arg Glu Pro Phe Trp Asn
 370 375 380

45

63

TTC GTC GGC ACC CGG ACG
 Phe Val Gly Thr Arg Thr
 245

738

5

(2) INFORMATIONS POUR LA SEQ ID NO: 21:

(i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 246 acides aminés
 (B) TYPE: acide aminé
 (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: protéine

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 21:

15

Met Arg Ala Asp Thr Glu Pro Thr Thr Gly Tyr Glu Asp Glu Phe Ala
 1 5 10 15

Glu Ile Tyr Asp Ala Val Tyr Arg Gly Arg Gly Lys Asp Tyr Ala Gly
 20 20 25 30

Glu Ala Lys Asp Val Ala Asp Leu Val Arg Asp Arg Val Pro Asp Ala
 35 40 45

25 Ser Ser Leu Leu Asp Val Ala Cys Gly Thr Gly Ala His Leu Arg His
 50 55 60

Phe Ala Thr Leu Phe Asp Asp Ala Arg Gly Leu Glu Leu Ser Ala Ser
 65 70 75 80

30

Met Leu Asp Ile Ala Arg Ser Arg Met Pro Gly Val Pro Leu His Gln
 85 90 95

Gly Asp Met Arg Ser Phe Asp Leu Gly Pro Arg Val Ser Ala Val Thr
 35 100 105 110

Cys Met Phe Ser Ser Val Gly His Leu Ala Thr Thr Ala Glu Leu Asp
 115 120 125

40 Ala Thr Leu Arg Cys Phe Ala Arg His Thr Arg Pro Gly Gly Val Ala
 130 135 140

Val Ile Glu Pro Trp Trp Phe Pro Glu Thr Phe Thr Asp Gly Tyr Val
 145 150 155 160

45

64

Ala Gly Asp Ile Val Arg Val Asp Gly Arg Thr Ile Ser Arg Val Ser
165 170 175

His Ser Val Arg Asp Gly Gly Ala Thr Arg Met Glu Ile His Tyr Val
5 180 185 190

Ile Ala Asp Ala Glu His Gly Pro Arg His Leu Val Glu His His Arg
195 200 205

10 Ile Thr Leu Phe Pro Arg His Ala Tyr Thr Ala Ala Tyr Glu Lys Ala
210 215 220

Gly Tyr Thr Val Glu Tyr Leu Asp Gly Gly Pro Ser Gly Arg Gly Leu
225 230 235 240

15

Phe Val Gly Thr Arg Thr
245

20 (2) INFORMATIONS POUR LA SEQ ID NO: 22:

(i) CARACTERISTIQUES DE LA SEQUENCE:

- 25 (A) LONGUEUR: 19 paires de bases
(B) TYPE: nucléotide
(C) NOMBRE DE BRINS: simple
(D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: Autre acide nucléique

- 30 (A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 22:

TCCTCGATGG AGACCTGCC

19

35

(2) INFORMATIONS POUR LA SEQ ID NO: 23:

(i) CARACTERISTIQUES DE LA SEQUENCE:

- 40 (A) LONGUEUR: 19 paires de bases
(B) TYPE: nucléotide
(C) NOMBRE DE BRINS: simple
(D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: Autre acide nucléique

- 45 (A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

Asp Asn
385

(2) INFORMATIONS POUR LA SEQ ID NO: 20:

5

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 738 paires de bases
(B) TYPE: nucléotide
(C) NOMBRE DE BRINS: double
10 (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN (génomique)

(vi) ORIGINE:

- 15 (A) ORGANISME: Streptomyces antibioticus

(ix) CARACTERISTIQUE:

- (A) NOM/CLE: CDS
(B) EMPLACEMENT:1..738
20 (D) AUTRES INFORMATIONS:/gene= "oleM"
/note= "SEQ ID NO 15 DE 3992 A 4729"

(ix) CARACTERISTIQUE:

- (A) NOM/CLE: mat_peptide
25 (B) EMPLACEMENT:1

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 20:

| | | |
|----|---|-----|
| 30 | ATG CGG GCT GAC ACG GAG CCG ACC ACC GGG TAC GAG GAC GAG TTC GCC | 48 |
| | Met Arg Ala Asp Thr Glu Pro Thr Thr Gly Tyr Glu Asp Glu Phe Ala | |
| | 1 5 10 15 | |
| | GAG ATC TAC GAC GCC GTG TAC CGG GGC CGG GGC AAG GAC TAC GCC GGC | 96 |
| 35 | Glu Ile Tyr Asp Ala Val Tyr Arg Gly Arg Gly Lys Asp Tyr Ala Gly | |
| | 20 25 30 | |
| | GAG GCG AAG GAC GTG GCG GAC CTC GTG CGC GAC CGG GTG CCG GAC GCG | 144 |
| | Glu Ala Lys Asp Val Ala Asp Leu Val Arg Asp Arg Val Pro Asp Ala | |
| 40 | 35 40 45 | |
| | TCC TCC CTC CTG GAC GTG GCC TGC GGC ACG GGC GCG CAC CTG CGG CAC | 192 |
| | Ser Ser Leu Leu Asp Val Ala Cys Gly Thr Gly Ala His Leu Arg His | |
| | 50 55 60 | |
| 45 | | |

62

| | | |
|----|---|-----|
| | TTC GCC ACG CTC TTC GAC GAC GCC CGC GGT CTC GAA CTG TCC GCG AGC | 240 |
| | Phe Ala Thr Leu Phe Asp Asp Ala Arg Gly Leu Glu Leu Ser Ala Ser | |
| | 65 70 75 80 | |
| 5 | ATG CTG GAC ATC GCC CGC TCC CGC ATG CCG GGC GTG CCG CTG CAC CAA | 288 |
| | Met Leu Asp Ile Ala Arg Ser Arg Met Pro Gly Val Pro Leu His Gln | |
| | 85 90 95 | |
| | GGG GAC ATG CGA TCC TTC GAC CTG GGG CCA CGC GTC TCC GCG GTC ACC | 336 |
| 10 | Gly Asp Met Arg Ser Phe Asp Leu Gly Pro Arg Val Ser Ala Val Thr | |
| | 100 105 110 | |
| | TGC ATG TTC AGC TCC GTC GGC CAC CTG GCC ACC ACC GCC GAA CTC GAC | 384 |
| | Cys Met Phe Ser Ser Val Gly His Leu Ala Thr Thr Ala Glu Leu Asp | |
| 15 | 115 120 125 | |
| | GCG ACG CTG CGG TGC TTC GCC CGG CAC ACC CGG CCC GGC GGC GTG GCC | 432 |
| | Ala Thr Leu Arg Cys Phe Ala Arg His Thr Arg Pro Gly Gly Val Ala | |
| | 130 135 140 | |
| 20 | GTC ATC GAA CCG TGG TGG TTC CCG GAG ACC TTC ACC GAC GGC TAC GTG | 480 |
| | Val Ile Glu Pro Trp Trp Phe Pro Glu Thr Phe Thr Asp Gly Tyr Val | |
| | 145 150 155 160 | |
| 25 | GCG GGT GAC ATC GTA CGC GTC GAC GGC CGG ACC ATC TCC CGG GTG TCC | 528 |
| | Ala Gly Asp Ile Val Arg Val Asp Gly Arg Thr Ile Ser Arg Val Ser | |
| | 165 170 175 | |
| | CAC TCG GTA CGG GAC GGC GGC GCC ACC CGC ATG GAG ATC CAC TAC GTG | 576 |
| 30 | His Ser Val Arg Asp Gly Gly Ala Thr Arg Met Glu Ile His Tyr Val | |
| | 180 185 190 | |
| | ATC GCC GAC GCC GAG CAC GGT CCC CGG CAC CTG GTC GAG CAC CAC CGC | 624 |
| | Ile Ala Asp Ala Glu His Gly Pro Arg His Leu Val Glu His His Arg | |
| 35 | 195 200 205 | |
| | ATC ACG CTG TTC CCG CGG CAT GCG TAC ACG GCC GCG TAC GAG AAG GCG | 672 |
| | Ile Thr Leu Phe Pro Arg His Ala Tyr Thr Ala Ala Tyr Glu Lys Ala | |
| | 210 215 220 | |
| 40 | GGC TAC ACC GTC GAG TAC CTC GAC GGC GGG CCC TCG GGC CGG GGG CTG | 720 |
| | Gly Tyr Thr Val Glu Tyr Leu Asp Gly Gly Pro Ser Gly Arg Gly Leu | |
| | 225 230 235 240 | |

45

65

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 23:

GAGACCATGC CCAGGGAGT

19

5 (2) INFORMATIONS POUR LA SEQ ID NO: 24:

(i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 19 paires de bases

(B) TYPE: nucléotide

10 (C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: Autre acide nucléique

15 (A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 24:

TCTGGGAGCC GCTCACCTT

19

20

(2) INFORMATIONS POUR LA SEQ ID NO: 25:

(i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 19 paires de bases

25 (B) TYPE: nucléotide

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: Autre acide nucléique

30 (A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 25:

35 GACGAGGCCG AAGAGGTGG

19

(2) INFORMATIONS POUR LA SEQ ID NO: 26:

(i) CARACTERISTIQUES DE LA SEQUENCE:

40 (A) LONGUEUR: 19 paires de bases

(B) TYPE: nucléotide

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

45 (ii) TYPE DE MOLECULE: Autre acide nucléique

(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 26:

GCACACCGGA ATGGATGCG

19

5

(2) INFORMATIONS POUR LA SEQ ID NO: 27:

(i) CARACTERISTIQUES DE LA SEQUENCE:

- 10 (A) LONGUEUR: 19 paires de bases
(B) TYPE: nucléotide
(C) NOMBRE DE BRINS: simple
(D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: Autre acide nucléique

- 15 (A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 27:

20 CCGTCGAGCT CTGAGGTAA

19

(2) INFORMATIONS POUR LA SEQ ID NO: 28:

(i) CARACTERISTIQUES DE LA SEQUENCE:

- 25 (A) LONGUEUR: 19 paires de bases
(B) TYPE: nucléotide
(C) NOMBRE DE BRINS: simple
(D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: Autre acide nucléique

- 30 (A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 28:

35

GCCCCGAGCCG CACGTGCGT

19

(2) INFORMATIONS POUR LA SEQ ID NO: 29:

(i) CARACTERISTIQUES DE LA SEQUENCE:

- 40 (A) LONGUEUR: 20 paires de bases
(B) TYPE: nucléotide
(C) NOMBRE DE BRINS: simple
(D) CONFIGURATION: linéaire

45

(ii) TYPE DE MOLECULE: Autre acide nucléique

(D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: Autre acide nucléique

(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

5

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 35:

CGCTCCAGGT GCAATGCCGG GTGCAGGC

28

10

(2) INFORMATIONS POUR LA SEQ ID NO: 36:

(i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 22 paires de bases

15

(B) TYPE: nucléotide

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: Autre acide nucléique

20

(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 36:

25 GATCACGCTC TTCGAGCGGC AG

22

(2) INFORMATIONS POUR LA SEQ ID NO: 37:

(i) CARACTERISTIQUES DE LA SEQUENCE:

30

(A) LONGUEUR: 21 paires de bases

(B) TYPE: nucléotide

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

35

(ii) TYPE DE MOLECULE: Autre acide nucléique

(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 37:

40

GAACTCGGTG GAGTCGATGT C

21

(2) INFORMATIONS POUR LA SEQ ID NO: 38:

45

(i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 21 paires de bases

70

- (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

- 5 (ii) TYPE DE MOLECULE: Autre acide nucléique
(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 38:

10

GTTGTCGATC AAGACCGCA C

21

(2) INFORMATIONS POUR LA SEQ ID NO: 39:

- 15 (i) CARACTERISTIQUES DE LA SEQUENCE:
(A) LONGUEUR: 22 paires de bases
(B) TYPE: nucléotide
(C) NOMBRE DE BRINS: simple
(D) CONFIGURATION: linéaire

20

- (ii) TYPE DE MOLECULE: Autre acide nucléique
(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

25 (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 39:

CATCGTCAAG GAGTTCGACG GT

22

(2) INFORMATIONS POUR LA SEQ ID NO: 40:

30

- (i) CARACTERISTIQUES DE LA SEQUENCE:
(A) LONGUEUR: 25 paires de bases
(B) TYPE: nucléotide
(C) NOMBRE DE BRINS: simple
35 (D) CONFIGURATION: linéaire

- (ii) TYPE DE MOLECULE: Autre acide nucléique
(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

40

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 40:

TGCGCAGGTC CATGTTACAC ACGTT

25

45 (2) INFORMATIONS POUR LA SEQ ID NO: 41:

(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 29:

5

TGCACGCGCT GCTGCCGACC

20

(2) INFORMATIONS POUR LA SEQ ID NO: 30:

10 (i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 20 paires de bases

(B) TYPE: nucléotide

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

15

(ii) TYPE DE MOLECULE: Autre acide nucléique

(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

20 (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 30:

TTGGCGAAGT CGACCAGGTC

20

(2) INFORMATIONS POUR LA SEQ ID NO: 31:

25

(i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 23 paires de bases

(B) TYPE: nucléotide

(C) NOMBRE DE BRINS: simple

30

(D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: Autre acide nucléique

(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

35

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 31:

GCCGCTCGGC ACGGTGAACT TCA

23

40 (2) INFORMATIONS POUR LA SEQ ID NO: 32:

(i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 24 paires de bases

(B) TYPE: nucléotide

45

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

68

- (ii) TYPE DE MOLECULE: Autre acide nucléique
(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

5 (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 32:

ATGCGCGTCG TCTTCTCCTC CATG

24

(2) INFORMATIONS POUR LA SEQ ID NO: 33:

10

- (i) CARACTERISTIQUES DE LA SEQUENCE:
(A) LONGUEUR: 21 paires de bases
(B) TYPE: nucléotide
(C) NOMBRE DE BRINS: simple
15 (D) CONFIGURATION: linéaire

- (ii) TYPE DE MOLECULE: Autre acide nucléique
(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

20

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 33:

TCATCGTGGT TCTCTCCTTC C

21

25 (2) INFORMATIONS POUR LA SEQ ID NO: 34:

- (i) CARACTERISTIQUES DE LA SEQUENCE:
(A) LONGUEUR: 23 paires de bases
(B) TYPE: nucléotide
30 (C) NOMBRE DE BRINS: simple
(D) CONFIGURATION: linéaire

- (ii) TYPE DE MOLECULE: Autre acide nucléique
(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

35

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 34:

GGAATTCATG ACCACGACCG ATC

23

40

(2) INFORMATIONS POUR LA SEQ ID NO: 35:

- (i) CARACTERISTIQUES DE LA SEQUENCE:
(A) LONGUEUR: 28 paires de bases
45 (B) TYPE: nucléotide
(C) NOMBRE DE BRINS: simple

71

- (i) CARACTERISTIQUES DE LA SEQUENCE:
 (A) LONGUEUR: 20 paires de bases
 (B) TYPE: nucléotide
 (C) NOMBRE DE BRINS: simple
5 (D) CONFIGURATION: linéaire
- (ii) TYPE DE MOLECULE: Autre acide nucléique
 (A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

10

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 41:

GCTACGCCCT GGAGAGCCTG

20

15 (2) INFORMATIONS POUR LA SEQ ID NO: 42:

- (i) CARACTERISTIQUES DE LA SEQUENCE:
 (A) LONGUEUR: 21 paires de bases
 (B) TYPE: nucléotide
20 (C) NOMBRE DE BRINS: simple
 (D) CONFIGURATION: linéaire
- (ii) TYPE DE MOLECULE: Autre acide nucléique
 (A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

25

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 42:

GTCGCGGTCTG GAGAGCACGA C

21

30

(2) INFORMATIONS POUR LA SEQ ID NO: 43:

- (i) CARACTERISTIQUES DE LA SEQUENCE:
 (A) LONGUEUR: 21 paires de bases
35 (B) TYPE: nucléotide
 (C) NOMBRE DE BRINS: simple
 (D) CONFIGURATION: linéaire
- (ii) TYPE DE MOLECULE: Autre acide nucléique
40 (A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 43:

45 GCCAGCTCGG CGACGTCCAT C

21

72

(2) INFORMATIONS POUR LA SEQ ID NO: 44:

(i) CARACTERISTIQUES DE LA SEQUENCE:

- 5 (A) LONGUEUR: 19 paires de bases
(B) TYPE: nucléotide
(C) NOMBRE DE BRINS: simple
(D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: Autre acide nucléique

- 10 (A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 44:

15 CGACGAGGTC GTGCATCAG

19

(2) INFORMATIONS POUR LA SEQ ID NO: 45:

(i) CARACTERISTIQUES DE LA SEQUENCE:

- 20 (A) LONGUEUR: 56 paires de bases
(B) TYPE: nucléotide
(C) NOMBRE DE BRINS: simple
(D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: Autre acide nucléique

- 25 (A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 45:

30

AATTGATCAA GGTGAACACG GTCATGCGCA GGATCCTCGA GCGGAAGTCC ATGGGG

56

(2) INFORMATIONS POUR LA SEQ ID NO: 46:

(i) CARACTERISTIQUES DE LA SEQUENCE:

- 35 (A) LONGUEUR: 56 paires de bases
(B) TYPE: nucléotide
(C) NOMBRE DE BRINS: simple
(D) CONFIGURATION: linéaire

40

(ii) TYPE DE MOLECULE: Autre acide nucléique

- (A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

45

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 46:

73

CCCCATGGAG TTCCGCTCGA GGATCCTGCG CATGACCGTG TTCACCTGA TCAATT

56

(2) INFORMATIONS POUR LA SEQ ID NO: 47:

5 (i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 32 paires de bases

(B) TYPE: nucléotide

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

10

(ii) TYPE DE MOLECULE: Autre acide nucléique

(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

15 (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 47:

AACTCGGTGG AGTCGATGTC GTCGCTGCGG AA

32

(2) INFORMATIONS POUR LA SEQ ID NO: 48:

20

(i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 27 paires de bases

(B) TYPE: nucléotide

(C) NOMBRE DE BRINS: simple

25

(D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: Autre acide nucléique

(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

30

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 48:

CAATATAGGA AGGATCAAGA GGTGAC

27

35 (2) INFORMATIONS POUR LA SEQ ID NO: 49:

(i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 39 paires de bases

(B) TYPE: nucléotide

40

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: Autre acide nucléique

(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

45

74

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 49:

TCCGGAGGTG TGCTGTCGGA CGGACTTGTC GGTCGGAAA

39

5 (2) INFORMATIONS POUR LA SEQ ID NO: 50:

(i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 33 paires de bases

(B) TYPE: nucléotide

10 (C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: Autre acide nucléique

(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

15

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 50:

20

AGGAGCACTA GTGCGGGTAC TGCTGACGTC CTT

33

(2) INFORMATIONS POUR LA SEQ ID NO: 51:

25 (i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 37 paires de bases

(B) TYPE: nucléotide

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

30

(ii) TYPE DE MOLECULE: Autre acide nucléique

(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

35 (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 51:

GGGGGATCCC ATATGCGGGT ACTGCTGACG TCCTTCG

37

(2) INFORMATIONS POUR LA SEQ ID NO: 52:

40

(i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 37 paires de bases

(B) TYPE: nucléotide

(C) NOMBRE DE BRINS: simple

45 (D) CONFIGURATION: linéaire

75

- (ii) TYPE DE MOLECULE: Autre acide nucléique
(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

5 (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 52:

GAAAAGATCT GCCGGCGTGG CGGCGCGTGA GTTCCTC

37

(2) INFORMATIONS POUR LA SEQ ID NO: 53:

10

- (i) CARACTERISTIQUES DE LA SEQUENCE:
(A) LONGUEUR: 27 paires de bases
(B) TYPE: nucléotide
(C) NOMBRE DE BRINS: simple
(D) CONFIGURATION: linéaire

15

- (ii) TYPE DE MOLECULE: Autre acide nucléique
(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

20

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 53:

AGCGGCTTGA TCGTGTGGA CCACTAC

27

25 (2) INFORMATIONS POUR LA SEQ ID NO: 54:

- (i) CARACTERISTIQUES DE LA SEQUENCE:
(A) LONGUEUR: 27 paires de bases
(B) TYPE: nucléotide
(C) NOMBRE DE BRINS: simple
(D) CONFIGURATION: linéaire

30

- (ii) TYPE DE MOLECULE: Autre acide nucléique
(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

35

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 54:

GGCCTATGTG GACTACGTGT TGAACGT

27

40

(2) INFORMATIONS POUR LA SEQ ID NO: 55:

- (i) CARACTERISTIQUES DE LA SEQUENCE:
(A) LONGUEUR: 31 paires de bases
(B) TYPE: nucléotide
(C) NOMBRE DE BRINS: simple

45

(D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: Autre acide nucléique

(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

5

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 55:

AACGCCTCGT CCTGCAGCGG AGACACGAAC A

31

10

(2) INFORMATIONS POUR LA SEQ ID NO: 56:

(i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 27 paires de bases

15

(B) TYPE: nucléotide

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: Autre acide nucléique

20

(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 56:

25 TTCGCTCCCC GATGAACACA ACTCGTA

27

(2) INFORMATIONS POUR LA SEQ ID NO: 57:

(i) CARACTERISTIQUES DE LA SEQUENCE:

30

(A) LONGUEUR: 35 paires de bases

(B) TYPE: nucléotide

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

35

(ii) TYPE DE MOLECULE: Autre acide nucléique

(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 57:

40

GAAGGAGATA TACATATGCG CGTCGTCTTC TCCTC

35

(2) INFORMATIONS POUR LA SEQ ID NO: 58:

45

(i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 32 paires de bases

77

- (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

- 5 (ii) TYPE DE MOLECULE: Autre acide nucléique
(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 58:

10

CGGGATCCTC ATCGTGGTTC TCTCCTTCCT GC

32

(2) INFORMATIONS POUR LA SEQ ID NO: 59:

- 15 (i) CARACTERISTIQUES DE LA SEQUENCE:
(A) LONGUEUR: 32 paires de bases
(B) TYPE: nucléotide
(C) NOMBRE DE BRINS: simple
(D) CONFIGURATION: linéaire

20

- (ii) TYPE DE MOLECULE: Autre acide nucléique
(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

25 (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 59:

CGGGTACCAT GCGCGTCGTC TTCTCCTCCA TG

32

(2) INFORMATIONS POUR LA SEQ ID NO: 60:

30

- (i) CARACTERISTIQUES DE LA SEQUENCE:
(A) LONGUEUR: 29 paires de bases
(B) TYPE: nucléotide
(C) NOMBRE DE BRINS: simple
(D) CONFIGURATION: linéaire

35

- (ii) TYPE DE MOLECULE: Autre acide nucléique
(A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE"

40

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 60:

CGGGTACCTC ATCGTGGTTC TCTCCTTCC

29

45 (2) INFORMATIONS POUR LA SEQ ID NO: 61:

(i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 13 acides aminés

(B) TYPE: acide aminé

(C) NOMBRE DE BRINS: simple

5 (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: peptide

(ix) CARACTERISTIQUE:

10 (A) NOM/CLE: Peptide

(B) EMPLACEMENT: 1..13

(D) AUTRES INFORMATIONS: /note= "SEQ ID NO 11 DE 38 A 50"

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 61:

15

Val Thr Gly Ala Gly Asp Gly Asp Ala Asp Val Gln Ala

1

5

10

INTERNATIONAL SEARCH REPORT

International Application No
PCT/FR 98/01593

A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/52 C12P19/62 C12Q1/68 C12N1/20 //(C12N1/20,
C12R1:01)

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12N C12P

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|------------|---|-----------------------|
| X | WO 97 23630 A (ABBOTT LAB) 3 July 1997 see figures 1,3,4A,4B,5 see page 2, line 37 - page 4, line 7 see page 10, line 3 - page 19, line 21 | 1-16, 25-32,41 |
| Y | see examples see claims | 17,18 |
| Y | SWAN D G ET AL: "Characterisation of a Streptomyces antibioticus gene encoding a type I polyketide synthase which has an unusual coding sequence." MOLECULAR AND GENERAL GENETICS, (1994 FEB) 242 (3) 358-62., XP002087278 see the whole document | 17,18 |
| | -/-- | |

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

Date of the actual completion of the international search

11 March 1999

Date of mailing of the international search report

13. 04. 1999

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Andres, S

INTERNATIONAL SEARCH REPORT

Int'l Application No
PCT/FR 98/01593

| C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT | | |
|--|---|-----------------------|
| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
| X | DONADIO S ET AL: "RECENT DEVELOPMENTS IN THE GENETICS OF ERYTHROMYCIN FORMATION" INDUSTRIAL MICROORGANISMS: BASIC AND APPLIED MOLECULAR GENETICS. ASM CONFERENCE ON THE GENETICS AND MOLECULAR BIOLOGY OF INDUSTRIAL MICROORGANISMS, 1992, pages 257-265, XP000607676 cited in the application see the whole document --- | 26-28, 32 |
| X | WO 97 06266 A (ABBOTT LAB) 20 February 1997 see example 1 see figure 2 --- | 10, 12, 13 |
| X | HAYDOCK S F ET AL: "CLONING AND SEQUENCE ANALYSIS OF GENES INVOLVE IN ERYTHROMYCIN BIOSYNTHESIS IN SACCHAROPOLYSPORA ERYTHRAEA: SEQUENCE SIMILARITIES BETWEEN ERYG AND A FAMILY OF S-ADENOSYLMETHIONINE-DEPENDENT METHYLTRANSFERASES" MOL. GEN. GENET, vol. 230, no. 1/02, November 1991, pages 120-128, XP002035888 cited in the application see figure 2 --- | 1-7 |
| X | WEBER J M ET AL: "ORGANIZATION OF A CLUSTER OF ERYTHROMYCIN GENES IN SACCHAROPOLYSPORA ERYTHRAEA" JOURNAL OF BACTERIOLOGY, vol. 172, no. 5, May 1990, pages 2372-2383, XP002035891 cited in the application see the whole document --- | 1, 2, 28, 29 |
| X | CAFFREY, P. ET AL.: "An acyl-carrier-protein-thioesterase domain from the 6-deoxyerythronolide B synthase of Saccharopolyspora erythraea" EUROPEAN JOURNAL OF BIOCHEMISTRY, vol. 195, February 1991, pages 823-830, XP002061258 see figures 2, 3 see page 827, left-hand column --- | 3, 5, 6 |
| X | WO 91 16334 A (ABBOTT LAB) 31 October 1991 see example 3 see figure 7 --- -/-- | 3, 5, 6 |



DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITE DE COOPERATION EN MATIÈRE DE BREVETS (PCT)

| | | |
|--|--|--|
| (51) Classification internationale des brevets ⁶ : C12N 15/52, C12P 19/62, C12Q 1/68, C12N 1/20 // (C12N 1/20, C12R 1:01) | A3 | (11) Numéro de publication internationale: WO 99/05283 (43) Date de publication internationale: 4 février 1999 (04.02.99) |
| (21) Numéro de la demande internationale: PCT/FR98/01593 (22) Date de dépôt international: 21 juillet 1998 (21.07.98) (30) Données relatives à la priorité: 97/09458 25 juillet 1997 (25.07.97) FR 98/07411 12 juin 1998 (12.06.98) FR (71) Déposant (pour tous les Etats désignés sauf US): HOECHST MARION ROUSSEL [FR/FR]; 1, terrasse Bellini, F-92800 Puteaux (FR). (72) Inventeurs; et (75) Inventeurs/Déposants (US seulement): FROMENTIN, Claude [FR/FR]; 16, rue de Flandres, F-75019 Paris (FR). MICHEL, Jean-Marc [FR/FR]; 22, rue des Domeliers, F-60200 Compiègne (FR). RAYNAL, Marie-Cécile [FR/FR]; 117, avenue de Choisy, F-75013 Paris (FR). SALAH-BEY, Khadidja [DZ/FR]; Appartement 2042, 100, boulevard Masséna, F-75013 Paris (FR). CORTES, Jesus [MX/GB]; 26 Cambanks, Union Lane, Cambridge CB4 1PZ (GB). GAISSER, Sabine [DE/GB]; 37 Gwydir Street, Cambridge CB1 2LG (GB). LEADLAY, Peter [GB/GB]; 17 Clarendon Road, Cambridge CB2 2BH (GB). MENDEZ, Carmen [ES/ES]; Calle Marcelino Fernandez 7, 2 ^o B, E-33010 Oviedo (ES). SALAS, Jose, A. [ES/ES]; Calle | Guillermo Estrada, 2-Bajo Izquierda, E-33060 Oviedo (ES). (74) Mandataire: VIEILLEFOSSE, Jean, Claude; Hoechst Marion Roussel, 102, route de Noisy, F-93235 Romainville Cedex (FR). (81) Etats désignés: BR, CA, JP, MX, TR, US, brevet européen (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Publiée <i>Avec rapport de recherche internationale.</i> <i>Avant l'expiration du délai prévu pour la modification des</i> <i>revendications, sera republiée si de telles modifications sont</i> <i>reçues.</i> (88) Date de publication du rapport de recherche internationale: 27 mai 1999 (27.05.99) | |
| (54) Title: BIOSYNTHESIS GENES AND TRANSFER OF 6-DESOXY-HEXOSES IN SACCHAROPOLYSPORA ERYTHRAEA AND IN STREPTOMYCES ANTIBIOTICUS AND THEIR USE (54) Titre: GENES DE BIOSYNTHESE ET DE TRANSFERT DES 6-DESOXYHEXOSES CHEZ SACCHAROPOLYSPORA ERYTHRAEA ET CHEZ STREPTOMYCES ANTIBIOTICUS ET LEUR UTILISATION (57) Abstract <p>The invention concerns the isolated DNA sequence represented in figure 2 (SEQ ID No. 1) corresponding to the <i>eryG-eryAIII</i> region of the cluster of the erythromycin biosynthesis genes and the isolated DNA sequence represented in figure 3 (SEQ ID No. 6) corresponding to the <i>eryAI-eryK</i> region of the cluster of the erythromycin biosynthesis genes. The invention also concerns the isolated DNA sequence represented in figure 22 (SEQ ID No. 15 sequence) corresponding to a region of the oleandomycin biosynthesis genes (SEQ ID No. 15 sequence).</p> (57) Abrégé <p>L'invention a pour objet la séquence d'ADN isolée représentée à la figure 2 (SEQ ID No. 1) correspondant à la région <i>eryG-eryAIII</i> du cluster de gènes de la biosynthèse de l'érythromycine et la séquence d'ADN isolée représentée à la figure 3 (SEQ ID No. 6) correspondant à la région <i>eryAI-eryK</i> du cluster de gènes de la biosynthèse de l'érythromycine, et a pour objet la séquence d'ADN isolée représentée à la figure 22 (séquence de SEQ ID No. 15) correspondant à une région du cluster de gènes de la biosynthèse de l'oléandomycine (séquence de SEQ ID No. 15).</p> | | |

UNIQUEMENT A TITRE D'INFORMATION

Codes utilisés pour identifier les Etats parties au PCT, sur les pages de couverture des brochures publiant des demandes internationales en vertu du PCT.

| | | | | | | | |
|----|---------------------------|----|---|----|--|----|-----------------------|
| AL | Albanie | ES | Espagne | LS | Lesotho | SI | Slovénie |
| AM | Arménie | FI | Finlande | LT | Lituanie | SK | Slovaquie |
| AT | Autriche | FR | France | LU | Luxembourg | SN | Sénégal |
| AU | Australie | GA | Gabon | LV | Lettonie | SZ | Swaziland |
| AZ | Azerbaïdjan | GB | Royaume-Uni | MC | Monaco | TD | Tchad |
| BA | Bosnie-Herzégovine | GE | Géorgie | MD | République de Moldova | TG | Togo |
| BB | Barbade | GH | Ghana | MG | Madagascar | TJ | Tadjikistan |
| BE | Belgique | GN | Guinée | MK | Ex-République yougoslave de Macédoine | TM | Turkménistan |
| BF | Burkina Faso | GR | Grèce | ML | Mali | TR | Turquie |
| BG | Bulgarie | HU | Hongrie | MN | Mongolie | TT | Trinité-et-Tobago |
| BJ | Bénin | IE | Irlande | MR | Mauritanie | UA | Ukraine |
| BR | Brésil | IL | Israël | MW | Malawi | UG | Ouganda |
| BY | Bélarus | IS | Islande | MX | Mexique | US | Etats-Unis d'Amérique |
| CA | Canada | IT | Italie | NE | Niger | UZ | Ouzbékistan |
| CF | République centrafricaine | JP | Japon | NL | Pays-Bas | VN | Viet Nam |
| CG | Congo | KE | Kenya | NO | Norvège | YU | Yougoslavie |
| CH | Suisse | KG | Kirghizistan | NZ | Nouvelle-Zélande | ZW | Zimbabwe |
| CI | Côte d'Ivoire | KP | République populaire démocratique de Corée | PL | Pologne | | |
| CM | Cameroun | KR | République de Corée | PT | Portugal | | |
| CN | Chine | KZ | Kazakhstan | RO | Roumanie | | |
| CU | Cuba | LC | Sainte-Lucie | RU | Fédération de Russie | | |
| CZ | République tchèque | LI | Liechtenstein | SD | Soudan | | |
| DE | Allemagne | LK | Sri Lanka | SE | Suède | | |
| DK | Danemark | LR | Libéria | SG | Singapour | | |
| EE | Estonie | | | | | | |

INTERNATIONAL SEARCH REPORT

Int. Patent Application No.

PCT/FR 98/01593

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|------------|---|-----------------------|
| X | STASSI D ET AL: "IDENTIFICATION OF A SACCHAROPOLYSPORA ERYTHRAEA GENE REQUIRED FOR THE FINAL HYDROXYLATION STEP IN ERYTHROMYCIN BIOSYNTHESIS" JOURNAL OF BACTERIOLOGY, vol. 175, no. 1, January 1993, pages 182-189, XP000608396 cited in the application see figures 2,5 | 8,9 |
| X | --- LIU, H.-W. & THORSON, J.: "Pathways and mechanisms in the biogenesis of novel deoxysugars by bacteria" ANNUAL REVIEW OF MICROBIOLOGY, vol. 48, 1994, pages 223-256, XP002061259 cited in the application see page 229, line 5 - page 232 | 41 |
| A | see page 236, line 28 - page 237, line 9 see page 250 see figure 9 | 1,8, 25-32 |
| A | --- QUIROS, L. & SALAS, J.: "Biosynthesis of the macrolide oleandomycin by Streptomyces antibioticus" JOURNAL OF BIOLOGICAL CHEMISTRY, (1995 AUG 4) 270 (31) 18234-9., vol. 270, 4 August 1995, pages 18234-18239, XP002096256 see the whole document | 19,20 |
| A | --- MALPARTIDA F ET AL: "Homology between Streptomyces genes coding for synthesis of different polyketides used to clone antibiotic biosynthetic genes" NATURE, vol. 325, 26 February 1987, pages 818-821, XP002075972 see the whole document | 17,18 |
| A | --- KAO, C. ET AL.: "Engineered biosynthesis of a complete macrolactone in a heterologous host" SCIENCE, vol. 265, 22 July 1994, pages 509-512, XP002096257 see abstract see page 511, left-hand column, paragraph 2 - right-hand column, last line --- -/-- | 39,40 |

INTERNATIONAL SEARCH REPORT

International Application No

PCT/FR 98/01593

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|------------|---|-----------------------|
| P,X | SUMMERS R G ET AL: "Sequencing and mutagenesis of genes from the erythromycin biosynthetic gene cluster of Saccharopolyspora erythraea that are involved in L-mycarose and D-desosamine production" MICROBIOLOGY-UK, (OCT 1997) VOL. 143, PART 10, PP. 3251-3262., XP002061260 see the whole document | 1-16, 25-32,41 |
| P,X | --- SALAH-BEY K ET AL: "Targeted gene inactivation for the elucidation of deoxysugar biosynthesis in the erythromycin producer Saccharopolyspora erythraea." MOLECULAR AND GENERAL GENETICS, (1998 MAR) 257 (5) 542-53., XP002087277 see the whole document | 1-16, 25-32 |
| P,X | --- GAISSER S ET AL: "Analysis of seven genes from the eryA1-eryK region of the erythromycin biosynthetic gene cluster in Saccharopolyspora erythraea" MOLECULAR & GENERAL GENETICS, (OCT 1997) VOL. 256, NO. 3, PP. 239-251., XP002061261 see the whole document | 8-16, 25-32,41 |
| P,X | --- GAISSER, S. ET AL.: "Analysis of eryBI, eryBIII and eryBVII from the erythromycin biosynthetic gene cluster in Saccharopolyspora erythraea" MOLECULAR AND GENERAL GENETICS., vol. 258, April 1998, pages 78-88, XP002087279 see the whole document | 8-10,12, 13,15,28 |
| T | --- OLANO C ET AL: "Analysis of a Streptomyces antibioticus chromosomal region involved in oleandomycin biosynthesis, which encodes two glycosyltransferases responsible for glycosylation of the macrolactone ring." MOLECULAR AND GENERAL GENETICS, (1998 AUG) 259 (3) 299-308., XP002096258 see the whole document ----- | 19-24, 39,40 |

INTERNATIONAL SEARCH REPORT

International application No.
PCT/FR 98/01593

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see supplementary sheet

1. ☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☒ The additional search fees were accompanied by the applicant's protest.
☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/FR 98/01593

The International Searching Authority found several (groups of) inventions in the international application, namely:

1. Claims: 1-18, 25-38 and 41

DNA sequences coding for proteins involved in erythromycin biosynthesis in *Saccharopolyspora*, corresponding polypeptides, their uses for the synthesis of hybrid secondary metabolites or as hybridization probes, and modified strains in one or the other of said genes. dTDP-D desosamine and its salts.

2. Claims: 19-24 and 39-40

DNA sequence comprising genes coding for proteins involved in oleomycin biosynthesis in *Streptomyces antibioticus*, and their use in a method for preparing oleandomycin precursors.

INTERNATIONAL SEARCH REPORT

Information on patent family members

national Application No

PCT/FR 98/01593

| Patent document cited in search report | | Publication date | Patent family member(s) | Publication date |
|---|---|---------------------|----------------------------|---------------------|
| WO 9723630 | A | 03-07-1997 | EP 0874548 A | 04-11-1998 |
| WO 9706266 | A | 20-02-1997 | CA 2201481 A | 20-02-1997 |
| | | | EP 0783584 A | 16-07-1997 |
| | | | JP 10507087 T | 14-07-1998 |
| WO 9116334 | A | 31-10-1991 | US 5141926 A | 25-08-1992 |
| | | | CA 2080583 A | 19-10-1991 |
| | | | EP 0525083 A | 03-02-1993 |
| | | | JP 2587562 B | 05-03-1997 |
| | | | JP 5504890 T | 29-07-1993 |
| | | | KR 9608668 B | 28-06-1996 |
| | | | PT 97390 A | 31-01-1992 |

RAPPORT DE RECHERCHE INTERNATIONALE

C inde internationale No
PCT/FR 98/01593

| | | |
|--|--|--|
| A. CLASSEMENT DE L'OBJET DE LA DEMANDE CIB 6 C12N15/52 C12P19/62 C12Q1/68 C12N1/20 //(C12N1/20, C12R1:01) | | |
| Selon la classification internationale des brevets (CIB) ou à la fois selon la classification nationale et la CIB | | |
| B. DOMAINES SUR LESQUELS LA RECHERCHE A PORTE Documentation minimale consultée (système de classification suivi des symboles de classement) CIB 6 C12N C12P | | |
| Documentation consultée autre que la documentation minimale dans la mesure où ces documents relèvent des domaines sur lesquels a porté la recherche | | |
| Base de données électronique consultée au cours de la recherche internationale (nom de la base de données, et si réalisable, termes de recherche utilisés) | | |
| C. DOCUMENTS CONSIDERES COMME PERTINENTS | | |
| Catégorie * | Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents | no. des revendications visées |
| X | WO 97 23630 A (ABBOTT LAB) 3 juillet 1997 voir figures 1,3,4A,4B,5 voir page 2, ligne 37 - page 4, ligne 7 voir page 10, ligne 3 - page 19, ligne 21 | 1-16, 25-32,41 |
| Y | voir exemples voir revendications --- | 17,18 |
| Y | SWAN D G ET AL: "Characterisation of a Streptomyces antibioticus gene encoding a type I polyketide synthase which has an unusual coding sequence." MOLECULAR AND GENERAL GENETICS, (1994 FEB) 242 (3) 358-62., XP002087278 voir le document en entier --- | 17,18 |
| --- -/-- | | |
| <div style="display: flex; justify-content: space-between;"> <input checked="" type="checkbox"/> Voir la suite du cadre C pour la fin de la liste des documents <input checked="" type="checkbox"/> Les documents de familles de brevets sont indiqués en annexe </div> | | |
| <div style="display: flex;"> <div style="flex: 1;"> <p>* Catégories spéciales de documents cités:</p> <p>"A" document définissant l'état général de la technique, non considéré comme particulièrement pertinent</p> <p>"E" document antérieur, mais publié à la date de dépôt international ou après cette date</p> <p>"L" document pouvant jeter un doute sur une revendication de priorité ou cité pour déterminer la date de publication d'une autre citation ou pour une raison spéciale (telle qu'indiquée)</p> <p>"O" document se référant à une divulgation orale, à un usage, à une exposition ou tous autres moyens</p> <p>"P" document publié avant la date de dépôt international, mais postérieurement à la date de priorité revendiquée</p> </div> <div style="flex: 1;"> <p>"T" document ultérieur publié après la date de dépôt international ou la date de priorité et n'appartenant pas à l'état de la technique pertinent, mais cité pour comprendre le principe ou la théorie constituant la base de l'invention</p> <p>"X" document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme nouvelle ou comme impliquant une activité inventive par rapport au document considéré isolément</p> <p>"Y" document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme impliquant une activité inventive lorsque le document est associé à un ou plusieurs autres documents de même nature, cette combinaison étant évidente pour une personne du métier</p> <p>"Z" document qui fait partie de la même famille de brevets</p> </div> </div> | | |
| Date à laquelle la recherche internationale a été effectivement achevée <div style="text-align: center; font-weight: bold;">11 mars 1999</div> | | Date d'expédition du présent rapport de recherche internationale <div style="text-align: center; font-weight: bold;">13. 04. 1999</div> |
| Nom et adresse postale de l'administration chargée de la recherche internationale Office Européen des Brevets, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016 | | Fonctionnaire autorisé <div style="text-align: center; font-weight: bold;">Andres, S</div> |

Formulaire PCT/ISA/210 (deuxième édition) (juillet 1992)

RAPPORT DE RECHERCHE INTERNATIONALE

ande Internationale No
PCT/FR 98/01593

| C.(suite) DOCUMENTS CONSIDERES COMME PERTINENTS | | |
|---|---|-------------------------------|
| Categorie | Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents | no. des revendications visées |
| X | DONADIO S ET AL: "RECENT DEVELOPMENTS IN THE GENETICS OF ERYTHROMYCIN FORMATION" INDUSTRIAL MICROORGANISMS: BASIC AND APPLIED MOLECULAR GENETICS. ASM CONFERENCE ON THE GENETICS AND MOLECULAR BIOLOGY OF INDUSTRIAL MICROORGANISMS, 1992, pages 257-265, XP000607676 cité dans la demande voir le document en entier --- | 26-28,32 |
| X | WO 97 06266 A (ABBOTT LAB) 20 février 1997 voir exemple 1 voir figure 2 --- | 10,12,13 |
| X | HAYDOCK S F ET AL: "CLONING AND SEQUENCE ANALYSIS OF GENES INVOLVE IN ERYTHROMYCIN BIOSYNTHESIS IN SACCHAROPOLYSPORA ERYTHAEA: SEQUENCE SIMILARITIES BETWEEN ERYG AND A FAMILY OF S-ADENOSYLMETHIONINE-DEPENDENT METHYLTRANSFERASES" MOL. GEN. GENET, vol. 230, no. 1/02, novembre 1991, pages 120-128, XP002035888 cité dans la demande voir figure 2 --- | 1-7 |
| X | WEBER J M ET AL: "ORGANIZATION OF A CLUSTER OF ERYTHROMYCIN GENES IN SACCHAROPOLYSPORA ERYTHRAEA" JOURNAL OF BACTERIOLOGY, vol. 172, no. 5, mai 1990, pages 2372-2383, XP002035891 cité dans la demande voir le document en entier --- | 1,2,28, 29 |
| X | CAFFREY, P. ET AL.: "An acyl-carrier-protein-thioesterase domain from the 6-deoxyerythronolide B synthase of Saccharopolyspora erythraea" EUROPEAN JOURNAL OF BIOCHEMISTRY, vol. 195, février 1991, pages 823-830, XP002061258 voir figures 2,3 voir page 827, colonne de gauche --- | 3,5,6 |
| X | WO 91 16334 A (ABBOTT LAB) 31 octobre 1991 voir exemple 3 voir figure 7 --- | 3,5,6 |
| -/-- | | |

RAPPORT DE RECHERCHE INTERNATIONALE

ande internationale No
PCT/FR 98/01593

| C.(suite) DOCUMENTS CONSIDERES COMME PERTINENTS | | |
|---|---|-------------------------------|
| Catégorie | Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents | no. des revendications visées |
| X | STASSI D ET AL: "IDENTIFICATION OF A SACCHAROPOLYSPORA ERYTHRAEA GENE REQUIRED FOR THE FINAL HYDROXYLATION STEP IN ERYTHROMYCIN BIOSYNTHESIS" JOURNAL OF BACTERIOLOGY, vol. 175, no. 1, janvier 1993, pages 182-189, XP000608396 cité dans la demande voir figures 2,5 --- | 8,9 |
| X | LIU, H.-W. & THORSON, J.: "Pathways and mechanisms in the biogenesis of novel deoxysugars by bacteria" ANNUAL REVIEW OF MICROBIOLOGY, vol. 48, 1994, pages 223-256, XP002061259 cité dans la demande voir page 229, ligne 5 - page 232 --- | 41 |
| A | voir page 236, ligne 28 - page 237, ligne 9 voir page 250 voir figure 9 --- | 1,8, 25-32 |
| A | QUIROS, L. & SALAS, J.: "Biosynthesis of the macrolide oleandomycin by Streptomyces antibioticus" JOURNAL OF BIOLOGICAL CHEMISTRY, (1995 AUG 4) 270 (31) 18234-9., vol. 270, 4 août 1995, pages 18234-18239, XP002096256 voir le document en entier --- | 19,20 |
| A | MALPARTIDA F ET AL: "Homology between Streptomyces genes coding for synthesis of different polyketides used to clone antibiotic biosynthetic genes" NATURE, vol. 325, 26 février 1987, pages 818-821, XP002075972 voir le document en entier --- | 17,18 |
| A | KAO, C. ET AL.: "Engineered biosynthesis of a complete macrolactone in a heterologous host" SCIENCE, vol. 265, 22 juillet 1994, pages 509-512, XP002096257 voir abrégé voir page 511, colonne de gauche, alinéa 2 - colonne de droite, dernière ligne --- | 39,40 |
| | -/-- | |

RAPPORT DE RECHERCHE INTERNATIONALE

Recherche internationale No
PCT/FR 98/01593

| C.(suite) DOCUMENTS CONSIDERES COMME PERTINENTS | | |
|---|---|-------------------------------|
| Categorie | Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents | no. des revendications visées |
| P,X | SUMMERS R G ET AL: "Sequencing and mutagenesis of genes from the erythromycin biosynthetic gene cluster of <i>Saccharopolyspora erythraea</i> that are involved in L-mycarose and D-desosamine production" MICROBIOLOGY-UK, (OCT 1997) VOL. 143, PART 10, PP. 3251-3262., XP002061260 voir le document en entier --- | 1-16, 25-32,41 |
| P,X | SALAH-BEY K ET AL: "Targeted gene inactivation for the elucidation of deoxysugar biosynthesis in the erythromycin producer <i>Saccharopolyspora erythraea</i> ." MOLECULAR AND GENERAL GENETICS, (1998 MAR) 257 (5) 542-53., XP002087277 voir le document en entier --- | 1-16, 25-32 |
| P,X | GAISSER S ET AL: "Analysis of seven genes from the eryA1-eryK region of the erythromycin biosynthetic gene cluster in <i>Saccharopolyspora erythraea</i> " MOLECULAR & GENERAL GENETICS, (OCT 1997) VOL. 256, NO. 3, PP. 239-251., XP002061261 voir le document en entier --- | 8-16, 25-32,41 |
| P,X | GAISSER, S. ET AL.: "Analysis of eryBI, eryBIII and eryBVII from the erythromycin biosynthetic gene cluster in <i>Saccharopolyspora erythraea</i> " MOLECULAR AND GENERAL GENETICS., vol. 258, avril 1998, pages 78-88, XP002087279 voir le document en entier --- | 8-10,12, 13,15,28 |
| T | OLANO C ET AL: "Analysis of a <i>Streptomyces antibioticus</i> chromosomal region involved in oleandomycin biosynthesis, which encodes two glycosyltransferases responsible for glycosylation of the macrolactone ring." MOLECULAR AND GENERAL GENETICS, (1998 AUG) 259 (3) 299-308., XP002096258 voir le document en entier ----- | 19-24, 39,40 |

RAPPORT DE RECHERCHE INTERNATIONALE

emande internationale n°
PCT/FR 98/01593

Cadre I Observations - lorsqu'il a été estimé que certaines revendications ne pouvaient pas faire l'objet d'une recherche (suite du point 1 de la première feuille)

Conformément à l'article 17.2)a), certaines revendications n'ont pas fait l'objet d'une recherche pour les motifs suivants:

1. ☐ Les revendications n^{os} se rapportent à un objet à l'égard duquel l'administration n'est pas tenue de procéder à la recherche, à savoir:
2. ☐ Les revendications n^{os} se rapportent à des parties de la demande internationale qui ne remplissent pas suffisamment les conditions prescrites pour qu'une recherche significative puisse être effectuée, en particulier:
3. ☐ Les revendications n^{os} sont des revendications dépendantes et ne sont pas rédigées conformément aux dispositions de la deuxième et de la troisième phrases de la règle 6.4.a).

Cadre II Observations - lorsqu'il y a absence d'unité de l'invention (suite du point 2 de la première feuille)

L'administration chargée de la recherche internationale a trouvé plusieurs inventions dans la demande internationale, à savoir:

voir feuille supplémentaire

1. ☒ Comme toutes les taxes additionnelles ont été payées dans les délais par le déposant, le présent rapport de recherche internationale porte sur toutes les revendications pouvant faire l'objet d'une recherche.
2. ☐ Comme toutes les recherches portant sur les revendications qui s'y prêtaient ont pu être effectuées sans effort particulier justifiant une taxe additionnelle, l'administration n'a sollicité le paiement d'aucune taxe de cette nature.
3. ☐ Comme une partie seulement des taxes additionnelles demandées a été payée dans les délais par le déposant, le présent rapport de recherche internationale ne porte que sur les revendications pour lesquelles les taxes ont été payées, à savoir les revendications n^{os}
4. ☐ Aucune taxe additionnelle demandée n'a été payée dans les délais par le déposant. En conséquence, le présent rapport de recherche internationale ne porte que sur l'invention mentionnée en premier lieu dans les revendications; elle est couverte par les revendications n^{os}

Remarque quant à la réserve

- ☒ Les taxes additionnelles étaient accompagnées d'une réserve de la part du déposant.
- ☐ Le paiement des taxes additionnelles n'était assorti d'aucune réserve.

SUITE DES RENSEIGNEMENTS INDICUES SUR PCT/ISA/ 210

L'administration chargée de la recherche internationale a trouvé plusieurs (groupe d') inventions dans la demande internationale, à savoir :

1. revendications: 1-18, 25-38 et 41

Séquences d'ADN codant pour des protéines impliquées dans la biosynthèse de l'érythromycine chez *Saccharopolyspora*, les polypeptides correspondant, leurs utilisations pour la synthèse de métabolites secondaires hybrides ou comme sondes d'hybridation, et des souches modifiées dans l'un ou l'autre de ces gènes. dTDP-D désosamine et ses sels.

2. revendications: 19-24 et 39-40

Séquence d'ADN comprenant des gènes codants pour des protéines impliquées dans la biosynthèse de l'oléomycine chez *Streptomyces antibioticus*, et leur utilisation dans un procédé pour la préparation de précurseurs de l'oléandomycine.

RAPPORT DE RECHERCHE INTERNATIONALE

Renseignements relatifs aux membres de familles de brevets

Demande internationale No

PCT/FR 98/01593

| Document brevet cité au rapport de recherche | Date de publication | Membre(s) de la famille de brevet(s) | Date de publication |
|---|------------------------|---|------------------------|
| WO 9723630 A | 03-07-1997 | EP 0874548 A | 04-11-1998 |
| WO 9706266 A | 20-02-1997 | CA 2201481 A | 20-02-1997 |
| | | EP 0783584 A | 16-07-1997 |
| | | JP 10507087 T | 14-07-1998 |
| WO 9116334 A | 31-10-1991 | US 5141926 A | 25-08-1992 |
| | | CA 2080583 A | 19-10-1991 |
| | | EP 0525083 A | 03-02-1993 |
| | | JP 2587562 B | 05-03-1997 |
| | | JP 5504890 T | 29-07-1993 |
| | | KR 9608668 B | 28-06-1996 |
| | | PT 97390 A | 31-01-1992 |

Formulaire PCT/ISA/210 (annexe familles de brevets) (juillet 1992)

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☒ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☒ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.